CORRELATION AND SCATTER IN STATISTICAL VARIABLES.

In the following lines the problem of linear correlation will be considered from the point of view of quadratic forms and linear transformations. The vector and matrix notation is found to be of great advantage, so it will be used extensively. I shall first define the notations used and state some classical facts from the algebra of matrices. The application to statistical variables of this algebraic tool will lead in a simple and most natural way to various results, some of wich are known, and others which are new so far as I am aware.

I. STATEMENT OF SOME FACTS FROM THE ALGEBRA OF MATRICES.

1. VECTORS.

A set of n quantities $x_1 \dots x_n$ is called a *vector* and denoted by a small heavy faced letter $\mathbf{x} = (x_1 \dots x_n)$. The quantities $x_1 \dots x_n$ are called the components of the vector. The i-th component of the vector $\mathbf{x}$ is also denoted $(\mathbf{x})_i$. In distinction to vectors, scalar quantities will be denoted by ordinary (not heavy faced) letters.

For $n = 2, 3$ $\mathbf{x}$ is represented geometrically by a directed straight line from origin to the point $(x_1 x_2)$ or $(x_1 x_2 x_3)$ respectively. For the sake of convenience the geometric language is used for any n. For our purpose it is immaterial if we think of $\mathbf{x}$ as representing the point $(x_1 \dots x_n)$ or the straight line from origin to this point. Sometimes it is convenient to speak of "the point $\mathbf{x}$" instead of "the vector $\mathbf{x}$".

The equation $\mathbf{x} = 0$ means that each component of $\mathbf{x}$ is zero. The product $c \cdot \mathbf{x}$, or shorter $c\mathbf{x}$, of the scalar c and the vector $\mathbf{x}$ is defined as the vector obtained by multiplying each component of $\mathbf{x}$ by c. The sum of two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as the vector whose i-th component is the sum of the i-th component of $\mathbf{x}$ and the i-th component of $\mathbf{y}$. Evidently $c\mathbf{x} = \mathbf{x}c$ and $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$. The product $\mathbf{x} \cdot \mathbf{y}$, or shorter $\mathbf{xy}$, is defined as the scalar $x_1 y_1 + \dots + x_n y_n$. Evidently $\mathbf{xy} = \mathbf{yx}$. The scalar $x = |\mathbf{x}| = +\sqrt{\mathbf{xx}}$ is called the modulus or the length of $\mathbf{x}$. The convexe angle $(\mathbf{xy})$ between the two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by $\cos(\mathbf{xy}) = \mathbf{xy}/xy$. Two vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to each other when, and only when $\mathbf{xy} = 0$. Two vectors $\mathbf{x}$ and $\mathbf{y}$ have the same (the opposite) direction when, and only when $\mathbf{xy} = +xy$ $(\mathbf{xy} = -xy)$. The vector $\mathbf{x}/x$ is the unit vector (i. e. the vector of length 1) in the direction $\mathbf{x}$.

The product $\mathbf{ax}$ can be looked upon as a linear form in $x_1 \cdots x_n$ with coefficients $a_1 \cdots a_n$. The equation $a_0 + \mathbf{ax} = 0$ represents a plane in $(x_1 \dots x_n)$ space. The unit vector $\mathbf{a}/a$ is the normal of this plane. The plane goes through the origin of $\mathbf{x}$ when, and only when $a_0 = 0$. The plane itself is defined indiscriminately either by the equation $a_0 + \mathbf{ax} = 0$ or by the equation $-a_0 - \mathbf{ax} = 0$. Fixing one of these two equations by convention means defining a positive and a negative side of the plane. If such a convention is made, the distance (measured perpendicularly) from a given point $\mathbf{x}'$ to the plane $a_0 + \mathbf{ax} = 0$ is equal to $(a_0 + \mathbf{ax}')/a$.

2. MATRICES.

A set of n^2 quantities a_{ij} $(i, j = 1, 2 \cdots n)$ is called a *matrix* and denoted by a heavy faced capital letter

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} \dots a_{1n} \\ \dots\dots\dots \\ a_{n1} \dots a_{nn} \end{pmatrix}$$

The quantities a_{ij} are called the elements of the matrix. The ij-th element of the matrix $\mathbf{A}$ is also denoted $(\mathbf{A})_{ij}$.

The diagonal containing the elements $a_{11} \dots a_{nn}$ is called the principal diagonal of $\mathbf{A}$. The elements in the principal diagonal are called the diagonal elements of $\mathbf{A}$. The vectors $\mathbf{a}_i = (a_{i1} \dots a_{in})$

are called the direct (vector) components, and the vectors $\grave{\mathbf{a}}_j = (a_{1j} \ldots a_{nj})$ the transposed (vector) components of $\mathbf{A}$.

The matrix obtained from $\mathbf{A}$ by interchanging rows and columns, is called the transposed of $\mathbf{A}$ and denoted $\grave{\mathbf{A}} = (\grave{a}_{ij})$, where $\grave{a}_{ij} = a_{ji}$. If $a_{ij} = a_{ji}$, $\mathbf{A}$ is called symmetric. A matrix with real elements is called a real matrix. $\mathbf{E} = \begin{pmatrix} 1 \ldots 0 \\ \cdots\cdots \\ 0 \ldots 1 \end{pmatrix}$ is called the unit matrix.

The equation $\mathbf{A} = 0$ means that each element of $\mathbf{A}$ is zero. The product $c \cdot \mathbf{A}$ or shorter $c\mathbf{A}$ of a scalar c and a matrix $\mathbf{A}$ is defined as the matrix obtained by multiplying each element of $\mathbf{A}$ by c. The sum of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as the matrix whose ij-th element is the sum of the ij-th element of $\mathbf{A}$ and the ij-th element of $\mathbf{B}$. Evidently $c\mathbf{A} = \mathbf{A}c$ and $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$.

The determinant value of the matrix $\mathbf{A}$ is called the modulus of $\mathbf{A}$ and denoted $A = |\mathbf{A}|$. The matrix (determinant) obtained by picking out q rows and q columns $(1 \leqq q \leqq n)$ from $\mathbf{A}$, is called a q rowed submatrix (q rowed minor) of $\mathbf{A}$; q is the order of the submatrix (minor). A submatrix (minor) whose diagonal elements are diagonal elements of $\mathbf{A}$, is called a principal submatrix (principal minor) of $\mathbf{A}$. $\mathbf{A}$ is said to be of rank ϱ $(1 \leqq \varrho \leqq n-1)$ if it contains at least one ϱ rowed minor which is different from zero, while all higher rowed minors vanish. $\mathbf{A}$ is said to be of rank n, or to be non singular if $A \neq 0$. Otherwise $\mathbf{A}$ is called singular. If $\mathbf{A} = 0$, $\mathbf{A}$ is said to be of rank 0. A necessary and sufficient condition for a *symmetric* matrix to be of rank ϱ, is that it contains at least one ϱ rowed *principal* minor, which is different from zero, while all higher rowed principal minors vanish.

A real symmetric matrix is called positive definite if all its principal minors (of all orders) are non negative. The case $\mathbf{A} = 0$ being however excluded. In this case $\mathbf{A}$ might de called zero definite. If a positive definite matrix is non singular, all its principal minors are positive, not zero. Hence the determinant value of a positive definite matrix which contains a vanishing principal minor, must be equal to zero. A necessary and sufficient condition for a real symmetric matrix $\mathbf{A}$ to be positive definite (respectively positive definite and non singular) is that all the n principal minors

$$a_{11}, \begin{vmatrix} a_{11} & a_{12} \\ \cdots & \cdots \\ a_{21} & a_{22} \end{vmatrix}, \ldots \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}$$

are non negative (respectively positive, not zero).

In a positive definite matrix all principal submatrices (of all orders) are positive definite. If $\mathbf{A}$ is positive definite, $-\mathbf{A}$ is called negative definite. If $\mathbf{A}$ is positive definite, $c\mathbf{A}$ is positive, negative or zero definite according as c is positive, negative or zero. In distinction to definite matrices, all other matrices are called indefinite.

The product of two matrices $\mathbf{AB}$ (taken in this order) is defined as the matrix whose ij-th element is $(\mathbf{AB})_{ij} = \sum_k a_{ik} b_{kj}$. This formula is analogous to the formula for multiplication of determinants. Hence the determinant of a product of two matrices is equal to the product of the determinants of the factors. A product of several matrices is associative but not commutative. A product of sums of matrices is distributive if account is taken of the difference between pre and post multiplication. The transposed of a product is equal to the product of the transposed taken in the reversed order. The rank of a product is never greater than the rank of any of the factors. The rank of a product of two matrices, one of which is non singular, is equal to the rank of the other matrix. If $\mathbf{A}$ has one of the properties of being definite, positive definite, definite and non singular etc., the matrix $\mathbf{CA\grave{C}}$ has the same property, provided $\mathbf{C}$ is non singular.

If $\grave{\mathbf{A}}$ is any matrix, $\grave{\mathbf{A}}\mathbf{A}$ is a symmetric matrix because it is a matrix which is equal to its transposed. A matrix $\mathbf{O}$ such that $\mathbf{O}\grave{\mathbf{O}} = \mathbf{E}$, that is $\sum_k o_{ik} o_{jk}$ is equal to 0 or 1 according as $i \neq j$ or $i = j$, is called an orthogonal matrix. The determinant of an orthogonal matrix must be equal to $+1$ or -1, for $|\mathbf{O}| \cdot |\grave{\mathbf{O}}| = |\mathbf{O}|^2 = |\mathbf{E}| = 1$.

p factors $\mathbf{A}$ is denoted $\mathbf{A}^p$ $(p \geqq 1)$, $\mathbf{A}^0$ is defined as equal to $\mathbf{E}$. Let $f(\lambda) = \sum_k a_k \lambda^k$ be a polynomial in λ. From the definitions laid down follows that $\mathbf{B} = f(\mathbf{A})$ is a uniquely determined matrix. $\mathbf{B}$ is called a polynomial matrix in the argument $\mathbf{A}$. If $f(\lambda) = g(\lambda) \cdot h(\lambda)$ identically in λ, g and h being polynomials, then $f(\mathbf{A}) = g(\mathbf{A}) \cdot h(\mathbf{A})$. A polynomial matrix in a symmetric argument is symmetric. If $\mathbf{A}$ is any matrix, there exists a polynomial in λ of degree $p \leqq n$, which vanishes if λ is replaced by $\mathbf{A}$. The polynomial $a(\lambda)$ of lowest degree such that $a(\mathbf{A}) = 0$, is called the *typical* polynomial for $\mathbf{A}$. By convention $a(\lambda)$ is determined so that the coefficient of the highest power of λ is equal to 1. Only one polynomial $a(\lambda)$ exists for a given $\mathbf{A}$.

The polynomial in λ, $A(\lambda) = |\mathbf{A} - \lambda\mathbf{E}|$ is called the *secular* (or characteristic) polynomial for $\mathbf{A}$, and the equation $A(\lambda) = 0$ is called

the secular equation for $\mathbf{A}$. The roots of the secular equation for $\mathbf{A}$ are called the characteristic numbers for $\mathbf{A}$. The secular equation for a real, symmetric matrix has only real roots, in particular it has the root $\lambda=0$ of multiplicity $n-\varrho$ when and only when the matrix is of rank ϱ. A real, symmetric matrix is positive (negative) definite when and only when all the roots of the secular equation are non negative (non positive).

A fundamental fact from the theory of matrices is that the typical polynomial $a(\lambda)$ is a divisor of the secular polynomial $A(\lambda)$. Hence the secular polynomial vanishes when λ is replaced by $\mathbf{A}$. Furthermore, the two polynomials $a(\lambda)$ and $A(\lambda)$ vanish for exactly the same values of λ. The only difference being that some of the zeros might occur whith a higher multiplicity in $A(\lambda)$ than in $a(\lambda)$.

The cofactor of a_{ij} in $\mathbf{A}$ is defined as $(-)^{i+j}$ times the minor of $\mathbf{A}$ which is obtained by omitting the row and column which intersect at a_{ij}, that is the i-th row and the j-th column. Let $\hat{a}_{ij}$ be the cofactor of a_{ji} in $\mathbf{A}$ (note the reversed order of the subscripts). The matrix $\hat{\mathbf{A}}=(\hat{a}_{ij})$ is called the *adjoint* of $\mathbf{A}$. If $\mathbf{A}$ is non singular, the matrix $\hat{\mathbf{A}}/A$ obtained by dividing each element of $\hat{\mathbf{A}}$ by $A=|\mathbf{A}|$, is called the *reciprocal* of $\mathbf{A}$ and denoted $\mathbf{A}^{-1}$. Any matrix has an adjoint, but only non singular matrices have a reciprocal. If $\mathbf{A}$ is of rank $\varrho=n$, $\varrho=n-1$, $\varrho\leqq n-2$, the adjoint $\hat{\mathbf{A}}$ is of rank n, 1 and 0 respectively. The adjoint (the reciprocal) of a product is equal to the product of the adjoints (of the reciprocals) taken in the reversed order. The adjoint (the reciprocal) of the transposed is equal to the transposed of the adjoint (of the reciprocal), so that the notation $\hat{\mathbf{A}}^{-1}$ is unambiguous. The adjoint of the transposed is denoted $\check{\mathbf{A}}$. The adjoint (the reciprocal) of symmetric matrix is symmetric. And the adjoint (the reciprocal) of a positive definite matrix is positive definite.

A simple application of the definition of a product shows that $\mathbf{AE}=\mathbf{EA}=\mathbf{A}$ and $\mathbf{AA}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{E}$. If $\mathbf{A}$ is non singular, the linear matrix equation $\mathbf{AX}=\mathbf{B}$ is therefore solved by premultiplication with $\mathbf{A}^{-1}$ wich gives $\mathbf{X}=\mathbf{A}^{-1}\mathbf{B}$. From this we infer $|\mathbf{A}^{-1}|=|\mathbf{A}|^{-1}$ and $|\hat{\mathbf{A}}|=A^n\,|\mathbf{A}^{-1}|=A^{n-1}$.

p factors $\mathbf{A}^{-1}$ is denoted $\mathbf{A}^{-p}$. Thus all integer powers of a matrix (with positive, negative or zero exponents) are defined, negative powers however being subject to the condition that the matrix shall be non singular. If p is any integer, $\mathbf{E}^p=\mathbf{E}$.

It is also possible to introduce fractional exponents. Here we

shall only consider square roots. $\sqrt{\mathbf{A}}=\mathbf{A}^{\frac{1}{2}}$ is defined as a matrix $\mathbf{B}$ such that $\mathbf{B}^2=\mathbf{A}$. By the fundamental theorem that the typical polynomial is a divisor of the secular polynomial, it is easy to prove that matrices $\mathbf{B}$ with the property $\mathbf{B}^2=\mathbf{A}$ always exist if $\mathbf{A}$ is non singular. If further $\mathbf{A}$ is positive definite, there even exist real matrices $\mathbf{B}$ such that $\mathbf{B}^2=\mathbf{A}$. In fact, the typical polynomial $a(\lambda)$ for $\mathbf{A}$ has a constant term different from zero, because $A(\lambda)$ has a constant term different from zero, namely $|\mathbf{A}|$. Therefore $a(\lambda)$ can be written in the form $a=(g^2-\lambda)/h$ where g and h are polynomials in λ. Since $a(\mathbf{A})=0$, $\mathbf{B}=g(\mathbf{A})$ is a matrix such that $\mathbf{B}^2=\mathbf{A}$. If further $\mathbf{A}$ is positive definite, all the zeros of $a(\lambda)$ are real and positive, and in this case $g(\lambda)$ might be chosen with real coefficients. The elements of $\mathbf{B}$ are consequently real.

A matrix $\mathbf{D}=\begin{pmatrix} d_1 \dots 0 \\ \dots\dots \\ 0 \dots d_n \end{pmatrix}$ which is of the same type as $\mathbf{E}$, having however $d_1 \dots d_n$ as its diagonal elements instead of $1, \dots 1$, is called a diagonal matrix. A diagonal matrix is evidently symmetric and its determinant value is equal to the product of its diagonal elements. A diagonal matrix is of rank ϱ if ϱ of its diagonal elements are different from zero. The adjoint (the reciprocal) of a diagonal matrix is a diagonal matrix. The product of two diagonal matrices $\mathbf{D}'$ and $\mathbf{D}''$ is a diagonal matrix whose diagonal elements are the products of the corresponding elements in $\mathbf{D}'$ and $\mathbf{D}''$. A diagonal matrix is raised to a (positive, negative or zero) power p, by raising each of its diagonal elements to the power p. If any matrix $\mathbf{A}$ is premultiplied (postmultiplied) by a diagonal matrix $\mathbf{D}$, the effect is to multiply each row (column) of $\mathbf{A}$ by the corresponding element of $\mathbf{D}$.

If $\mathbf{A}$ is any matrix whose diagonal elements are different from zero and of the same sign, the matrix obtained by dividing all the elements a_{ij} by $+\sqrt{a_{ii}a_{jj}}$, is called the diagonally normalized matrix of $\mathbf{A}$, or shorter the normalized of $\mathbf{A}$, The normalized of $\mathbf{A}$ can be expressed as $\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{D}=\begin{pmatrix} a_{11} \dots 0 \\ \dots\dots \\ 0 \dots a_{nn} \end{pmatrix}$, $\mathbf{D}$ is called the diagonal matrix of $\mathbf{A}$. In any normalized matrix the elements of the principal diagonal are all equal to 1. The normalized of $\mathbf{A}$ is equal to the normalized of $\mathbf{D}'\mathbf{A}\mathbf{D}'$, where $\mathbf{D}'$ is an arbitrary non singular diagonal matrix whose diagonal elements are all of the same sign.

3. MATRICES MULTIPLIED BY ONE VECTOR.

The product $\mathbf{Ax}$ is defined as the vector $\mathbf{y}$ whose i-th component is $y_i=(\mathbf{Ax})_i=\Sigma_k a_{ik}x_k$. The product $\mathbf{xA}$ is defined as the vector $\mathbf{y}$ whose j-th component is $y_j=(\mathbf{xA})_j=\Sigma_k x_k a_{kj}$. Evidently $\mathbf{xA}=\grave{\mathbf{A}}\mathbf{x}$ and $\mathbf{Ex}=\mathbf{x}$.

A product of several matrices and one vector is associative if the vector stands either before or after all the matrices. An expression like $\mathbf{ABCx}$ is therefore unambiguous. But $(\mathbf{Ax})\mathbf{B}$ will in general be different from $\mathbf{A}(\mathbf{xB})$. By introducing the transposed matrices it is always possible to carry the vector to one end of the product and thus make the product associative. For instance $(\mathbf{Ax})\mathbf{B}=\mathbf{x}\grave{\mathbf{A}}\mathbf{B}=\grave{\mathbf{B}}\mathbf{Ax}$.

The vector $\mathbf{y}=\mathbf{Ax}$ can be looked upon as representing n linear forms $y_i=\Sigma_j a_{ij}x_j$ If the components of $\mathbf{y}$ are given, while those of $\mathbf{x}$ are unknown, the equation $\mathbf{Ax}=\mathbf{y}$ may be looked upon as a system of linear equations. If $\mathbf{A}$ is non singular, the system is solved by premultiplication with $\mathbf{A}^{-1}$, which gives $\mathbf{x}=\mathbf{A}^{-1}\mathbf{y}$.

The equation $\mathbf{x}'=\mathbf{Ax}$ can be looked upon as representing a homogeneous linear transformation, whereby the set of variables $\mathbf{x}=(x_1 \ldots x_n)$ is replaced by the set $\mathbf{x}'=(x_1' \ldots x_n')$. $\mathbf{A}$ is called the matrix and A the modulus of the transformation. A transformation is called non singular, orthogonal etc. according as $\mathbf{A}$ has these properties.

If we have two sets of variables $\mathbf{x}$ and $\mathbf{y}$, and both sets are subject to the same transformation $\mathbf{x}'=\mathbf{Ax}$ and $\mathbf{y}'=\mathbf{Ay}$, $\mathbf{x}$ and $\mathbf{y}$ are called *cogredient* sets. If $\mathbf{x}'=\mathbf{Ax}$ while $\grave{\mathbf{A}}\mathbf{y}'=\mathbf{y}$, $\mathbf{x}$ and $\mathbf{y}$ are called *contragredient* sets.

4. MATRICES MULTIPLIED BY TWO VECTORS.

The notation $\mathbf{xAy}$ is unambiguous because $\mathbf{x}(\mathbf{Ay})$ and $(\mathbf{xA})\mathbf{y}$ according to the definitions laid down in the preceeding sections, both represent the same thing, namely a scalar, the bilinear form $\mathbf{xAy}=\Sigma_{ij}x_i a_{ij}y_j$. If $\mathbf{x}=\mathbf{y}$, we have the quadratic form $\mathbf{xAx}=\Sigma_{ij}x_i a_{ij}x_j$. Evidently $\mathbf{xAy}=\mathbf{y}\grave{\mathbf{A}}\mathbf{x}$ and $\mathbf{xEy}=\mathbf{xy}$.

A product of several matrices and two vectors is associative, if one vector stands before and the other after the string of matrices. The explicit formula for a product like $\mathbf{xABCy}$ is

$$\mathbf{xABCy}=\Sigma_{ij}\Sigma_{kh}x_i a_{ik}b_{kh}c_{hj}y_j$$

This is a bilinear form with the ij-th coefficient equal to $\Sigma_{kh}\, a_{ik}b_{kh}\, c_{hj}$. By introducing the transposed matrices it is always possible to carry one vector to each end of the product and thus make the product associative. For instance $(\mathbf{Ax})\cdot[(\mathbf{By})\mathbf{C}]=\mathbf{x}\grave{\mathbf{A}}\grave{\mathbf{C}}\mathbf{By}$.

One of the advantages of the notations and rules laid down in this and the preceeding sections, is that bilinear forms themselves and transformations performed on such forms may be considered from the same point of view. The notions of operand and operator thus become equivalent, much in the same way as multiplicand and multiplicator are equivalent notions in ordinary algebra. The following is a simple illustration. If we perform the transformations $\mathbf{x}=\mathbf{Px}'$ and $\mathbf{y}=\mathbf{Qy}'$ on the variables in the bilinear form $\mathbf{xAy}$, the form into which $\mathbf{xAy}$ is carried, is simply determined by introducing for $\mathbf{x}$ and $\mathbf{y}$, rearranging the factors and multiplying out the matrix product obtained: $\mathbf{xAy}=(\mathbf{Px}')\mathbf{A}(\mathbf{Qy}')=$ $=\mathbf{x}'(\grave{\mathbf{P}}\mathbf{AQ})\mathbf{y}'$. In particular, if $\mathbf{x}$ and $\mathbf{y}$ are cogredient, i. e, if the transformations are $\mathbf{x}=\mathbf{Px}'$ and $\mathbf{y}=\mathbf{Py}'$, we get $\mathbf{xAy}=\mathbf{x}'(\grave{\mathbf{P}}\mathbf{AP})\mathbf{y}'$. If $\mathbf{x}$ and $\mathbf{y}$ are contragredient, i. e. if the transformations are $\mathbf{x}=\grave{\mathbf{Q}}\mathbf{x}'$, $\mathbf{y}=\mathbf{Q}^{-1}\mathbf{y}'$, we have $\mathbf{xAy}=\mathbf{x}'(\mathbf{QAQ}^{-1})\mathbf{y}'$.

In the case of a quadratic form it does not restrict generality if we assume $\mathbf{A}$ to be symmetric. In the following this assumption is always to be understood when quadratic forms are discussed. The quadratic form $\mathbf{xAx}$ is called real if $\mathbf{A}$ is real.

A real quadratic form $\mathbf{xAx}$ is called definite if it keeps a constant sign for all possible real values of the variables (the form eventually vanishing for certain sets of values of the variables). A definite form is called positive (negative) definite if its constant sign is non negative (non positive). A definite form is called zero definite if it vanishes identically in $\mathbf{x}$. This is the case when and only when $\mathbf{A}=0$ (i. e. when the rank of $\mathbf{A}$ is $\varrho=0$). A definite form is called definite and non singular if it does not vanish for any other set of values of the variables than $\mathbf{x}=0$ (i. e. $x_1=\ldots=x_n=0$).

In distinction to definite forms all other real, quadratic forms are called indefinite. An indefinite form is therefore a real, quadratic form which for certain real values of the variables is positive, not zero, and for certain other real values of the variables is negative, not zero.

A quadratic form is definite, positive (negative) definite, definite and non singular etc. according as its matrix has these properties. Criteria for these cases have been given in Section 2. In parti-

cular, if $\mathbf{A}$ is a positive definite and non singular matrix, the form $\mathbf{xAx}$ is positive, not zero, for all real values of the variables, except $\mathbf{x}=0$.

A classical fact from the theory of quadratic forms is that if $\mathbf{xAx}$ is any real, quadric form, there exists a real and non singular transformation $\mathbf{x}=\mathbf{Px}'$ which carries the given form over into a sum of squares. That is

$$\mathbf{xAx}=\mathbf{x}'(\grave{\mathbf{P}}\mathbf{AP})\mathbf{x}'=\mathbf{Dx}'$$

where $\mathbf{D}=\grave{\mathbf{P}}\mathbf{AP}$ is a diagonal matrix. There even exists an infinity of such transformations $\mathbf{P}$. In particular there exists a real orthogonal $\mathbf{P}$. Whatever the particular form of $\mathbf{P}$ might be, the number of diagonal elements in $\mathbf{D}$ which are positive, negative and zero respectively, is always the same and determined by the nature of $\mathbf{A}$. In particular the number of diagonal elements in $\mathbf{D}$ which are equal to zero is $n-\varrho$, where ϱ is the rank of $\mathbf{A}$. This simply follows from the fact that if $\mathbf{P}$ is non singular, $\mathbf{A}$ and $\mathbf{D}$ must have the same rank. A diagonal element in $\mathbf{D}$ might be called effective or ineffective according as it is different from zero or equal to zero. All the diagonal elements in $\mathbf{D}$ are non negative (non positive) when and only when $\mathbf{A}$ is positive (negative) definite. The number of diagonal elements in $\mathbf{D}$ which are positive, not zero, is called the index of $\mathbf{A}$. A positive definite matrix is therefore a matrix whose index is equal to its rank. A consequence of the facts stated above, is that if $\mathbf{C}$ denotes an arbitrary non singular matrix, the matrix $\grave{\mathbf{C}}\mathbf{AC}$ has the same rank and the same index as the matrix $\mathbf{A}$. In fact, putting $\mathbf{Q}=\mathbf{C}^{-1}\mathbf{P}$ and $\mathbf{B}=\grave{\mathbf{C}}\mathbf{AC}$, we have $\grave{\mathbf{Q}}\mathbf{BQ}=\mathbf{D}$.

5. MAXIMA AND MINIMA OF A FUNCTION OF n VARIABLES.

The essential facts regarding maxima and minima of a function of several variables can be stated in a very simple way by using the vector and matrix notation.

Let $f(x_1 \ldots x_n)$ or shorter $f(\mathbf{x})$ be a real function of the set of n independent, real variables $\mathbf{x}=(x_1 \ldots x_n)$. We assume that $f(\mathbf{x})$ in a certain domain of $(x_1 \ldots x_n)$ space has continuous partial derivatives up to the highest order involved in the following argument.

Let $f_i=\frac{\partial f}{\partial x_i}$, $f_{ij}=\frac{\partial^2 f}{\partial x_i \partial x_j}$ and put $\mathbf{f}=(f_1 \ldots f_n)$ $\mathbf{F}=\begin{pmatrix} f_{11} \ldots f_{1n} \\ \ldots\ldots\ldots \\ f_{n1} \ldots f_{nn} \end{pmatrix}$.

In order to indicate that the components of $\mathbf{f}$ and the elements of $\mathbf{F}$ are functions of the variables in the set $\mathbf{x}$, we use the notation

$$\mathbf{f}=\mathbf{f}(\mathbf{x}) \qquad \mathbf{F}=\mathbf{F}(\mathbf{x}).$$

Now, let $\mathbf{x}$ be a fixed point, and consider the distribution of the values of f in the vicinity of $\mathbf{x}$. Let $\mathbf{h}=(h_1 \ldots h_n)$ where h_i is a small increment to the variable x_i. The vector $\mathbf{h}$ represents the total displacement from $\mathbf{x}$. The *Taylor* expansion of f in terms of the increments h_i, can be written

$$f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})=\frac{\mathbf{f}(\mathbf{x})\cdot\mathbf{h}}{1!}+\frac{\mathbf{h}\cdot\mathbf{F}(\mathbf{x})\cdot\mathbf{h}}{2!}+\ldots\ldots$$

The analogy between this formula and the formula for a function of a single variable is obvious.

Now, if $\mathbf{x}$ shall be either a maximum point or a minimum point for f, it is necessary that

$$\mathbf{f}(\mathbf{x})=0.$$

In fact, if one (or more) of the components of $\mathbf{f}=\mathbf{f}(\mathbf{x})$ were $\neq 0$, we could choose $\mathbf{h}$ in such ways that the term $\mathbf{fh}$ (which for a sufficiently small $|\mathbf{h}|$ determines the sign of the right hand side of the expansion) assumed first a positive and next a negative value. The difference $f(\mathbf{x}+\mathbf{h})-f(\mathbf{x})$ would consequently not keep a constant sign in the vicinity of $\mathbf{x}$. The eventually existing points where f attains a maximum or a minimum, must therefore be found among the solutions $\mathbf{x}$ of $\mathbf{f}(\mathbf{x})=0$.

If any such solution shall actually furnish a maximum or a minimum depends on the nature of the matrix $\mathbf{F}=\mathbf{F}(\mathbf{x})$. This matrix namely determines the distribution of the values of $\mathbf{hFh}$. And when $\mathbf{f}=0$, $\mathbf{hFh}$ will be the sign determining term in the right hand side of the expansion. The following cases are possible.

(1) $\mathbf{F}$ *is zero definite.* In this case the rank of $\mathbf{F}$ is $\varrho=0$, i. e. each element of $\mathbf{F}$ is zero. The quadratic form $\mathbf{hFh}$ therefore vanishes identically in $\mathbf{h}$. This is the case which can not be decided upon by the properties of $\mathbf{F}$ alone, but where the higher partial derivatives of f must be taken into account.

(2) $\mathbf{F}$ *is negative definite.* In this case the quadratic form $\mathbf{hFh}$ can never assume a positive value, so long as $\mathbf{h}$ is real. We therefore immediately infer that the point $\mathbf{x}$ now is a maximum point in the broad sense that there are no points in the vicinity

of $\mathbf{x}$ where f assumes a greater value than in $\mathbf{x}$. In a narrower sense however, $\mathbf{x}$ might not be a maximum point. We have to distinguish between the following two cases: (a) There exist points in the vicinity of $\mathbf{x}$ where f assumes a value as large as $f(\mathbf{x})$. In this case $\mathbf{x}$ is called an improper maximum point. (b) In all points in the vicinity of $\mathbf{x}$, the value of f is less than $f(\mathbf{x})$. In this case $\mathbf{x}$ is called a proper maximum point.

Criteria for these two cases are readily derived from $\mathbf{F}$. If $\mathbf{F}$ is of rank ϱ, there exists an $n-\varrho$ dimensional plane through $\mathbf{x}$, such that in this plane (in the vicinity of $\mathbf{x}$), f is constantly equal to $f(\mathbf{x})$. Anywhere else in the vicinity of $\mathbf{x}$ the value of f is less than $f(\mathbf{x})$. The vicinity of $\mathbf{x}$ is assumed infinitesimal of the second order. If infinitesimal increments of higher order than the second are taken into account, the $n-\varrho$ dimensional variety where f is constant, can in general not be looked upon as a plane.

To prove the above criterion let us performe a non singular real transformation $\mathbf{h}=\mathbf{Ph}'$ which carries the form $\mathbf{hFh}$ over into $\mathbf{hFh}=\mathbf{h}'\mathbf{Dh}'$, where $\mathbf{D}=\mathbf{PFP}$ is a diagonal matrix with $n-\varrho$ of its diagonal elements ineffective, i. e. equal to zero, and the remaining ϱ diagonal elements effective, i. e. different from zero. These ϱ effective diagonal elements of $\mathbf{D}$ must be negative since $\mathbf{F}$ is negative definite. In order that $\mathbf{hFh}$ shall be equal to zero it is therefore necessary and sufficient that the ϱ components of $\mathbf{h}'$ which correspond to effective diagonal elements in $\mathbf{D}$, are zero. Putting these ϱ components of $\mathbf{h}'$ equal to zero in the expression $\mathbf{h}=\mathbf{Ph}'$ we obtain an expression for $\mathbf{h}$ where there remain $n-\varrho$ arbitrary parameters. This expression represents an $n-\varrho$ dimensional plane in which f is constantly equal to $f(\mathbf{x})$ (a part from infinitesimal increments of higher order than the second). In any point outside this plane (and in the vicinity of $\mathbf{x}$) f must be less than $f(\mathbf{x})$ because any such point corresponds to a vector $\mathbf{h}'$ such that $\mathbf{h}'\mathbf{Dh}'$ is negative, not zero.

In particular we see that $\mathbf{x}$ is a proper maximum point when and only when $\varrho=n$, i. e. when and only when the matrix $\mathbf{F}$ is negative definite and non singular. If $\varrho=n-p$, the minimum might be called p-fold improper. If the highest part of a surface has the shape, not of a peak, but of an horizontal edge, any point on this edge represents an improper maximum point. In this case we have $n=2$ and $\varrho=1$. The straight line which, in a given point, is tangent to the curve obtained by projecting the edge

on to a horisontal plane, represents the $p=1$ dimensional "plane" where $f(\mathbf{x})$ is constant (a part from infinitesimal increments of higher order than the second).

(3) $\mathbf{F}$ *is positive definite.* In this case $\mathbf{x}$ is always a minimum point in the broad sense of the word. The minimum point is proper or improper according as $\mathbf{F}$ is non singular or singular. The discussion is perfectly similar to the discussion in the case (2).

(4) $\mathbf{F}$ *is indefinite.* In this case $\mathbf{x}$ is neither a maximum point nor a minimum point. In certain sections of the vicinity of $\mathbf{x}$ f is less than $f(\mathbf{x})$, and in certain other sections of the vicinity f is greater than $f(\mathbf{x})$. The quadratic form $\mathbf{hFh}$ will namely now be negative, not zero, for certain vectors $\mathbf{h}$, and positive, not zero, for certain other vectors $\mathbf{h}$.

The cases (1), (2) and (3) correspond to the cases where the second derivative of a function of a single variable is zero, negative and positive respectively. For a function of a single variable there exists no analogon to the case (4). A matrix consisting of a single element is namely always definite.

II. STATISTICAL VARIABLES.

1. THE MOMENT MATRIX.

Suppose there is made a certain number ω of statistical observations, each observation being characterized by the values of n quantitative attributes $z_1 \ldots z_n$. These quantities might be interpreted as the components of a vector $\mathbf{z}=(z_1 \ldots z_n)$. For the following analysis it is immateriel if the observations are thought of as ordered in time or simply as elements in a statistical population. We assume the observations ordered in time $\mathbf{z}(t)=(z_1(t) \ldots z_n(t))$ simply for the sake of convenience of expression. Thus summation over time Σ_t only means summation over all the ω observations. The variables measured from their means will be denoted $\mathbf{x}=(x_1 \ldots x_n)$, so that $\Sigma_t \mathbf{x}(t)=0$, that is $\Sigma_t x_i(t)=0$ $(i=1,2 \ldots n)$. The variables $\mathbf{z}$ and $\mathbf{x}$ here considered will be called *observational* variables as distinguished from certain symbolic variables which will be introduced later. The set $\mathbf{z}$, respectively the set $\mathbf{x}$, will be called an n dimensional (observational) set. Most of the time we shall consider the set $\mathbf{x}$ instead of $\mathbf{z}$. If each actual observation (i. e. for each value of t) is represented by a point in n

dimensional space, we obtain a swarm of observation points called the *scatter diagram.*

The product moments m_{ij} (taken about the means) are defined as the quantities $m_{ij}=\Sigma_t x_i(t)x_j(t)$. The symmetric matrix

$$\mathbf{M}=(m_{ij})=\begin{pmatrix} m_{11} \dots m_{1n} \\ \dots\dots\dots \\ m_{n1} \dots m_{nn} \end{pmatrix}$$

will be called the *moment matrix* of the set $\mathbf{x}$. Some of the most essential features of the statistical material at hand, are revealed by the properties of this matrix.

The i-th diagonal element m_{ii} of $\mathbf{M}$ is the sum square of the variable x_i, hence non negative; $\sigma_i=+\sqrt{m_{ii}/\omega}$ is the standard deviation of x_i. An observational variable (which is measured from its mean) is identically zero when and only when its standard deviation is zero. A variable might be called effective or ineffective according as its standard deviation is different from zero or equal to zero. If all the variables are effective, the set might be called an effective set. If no statement to the contrary is made, we shall assume the set considered to be an effective set. In this case the diagonal matrix of $\mathbf{M}$, namely

$$\mathbf{D}=\begin{pmatrix} m_{11} \dots\dots 0 \\ \dots\dots\dots\dots \\ 0 \dots\dots m_{nn} \end{pmatrix}$$

is non singular.

The set

$$\mathbf{y}=\sqrt{\omega}\mathbf{D}^{-\frac{1}{2}}\mathbf{x} \qquad \text{i. e. } y_i=x_i/\sigma_i \tag{1.1}$$

obtained from the effective set $\mathbf{x}$ by dividing each variable by its standard deviation, may be called the standardized set of $\mathbf{x}$, $\mathbf{y}$ are the *standard coordinates* of $\mathbf{x}$, the set $\mathbf{y}$ is obtained by standardizing the set $\mathbf{x}$.

The q rowed principal submatrix $\mathbf{M}_{q[\alpha \dots \gamma]}$ formed by the q rows $\alpha \dots \gamma$ and the q columns $\alpha \dots \gamma$ from $\mathbf{M}$ is the moment matrix of the subset $x_\alpha \dots x_\gamma$. The determinant values of the moment matrices will be denoted

$$M=|\mathbf{M}| \qquad M_{q[\alpha \dots \gamma]}=|\mathbf{M}_{q[\alpha \dots \gamma]}|$$

$\hat{\mathbf{M}}$ will be called the adjoint moment matrix.

I shall now consider some of the properties of $\mathbf{M}$. I shall first show that $\mathbf{M}$ is *positive definite.*

In fact $M_{q[\alpha \dots \gamma]}$ is a *Gram*-ian determinant. Hence we have the expression[1]

$$q!\,M_{q[\alpha \dots \gamma]}=\Sigma_{t_1 \dots t_q}\begin{vmatrix} x_\alpha(t_1) \dots x_\alpha(t_q) \\ \dots\dots\dots\dots \\ x_\gamma(t_1) \dots x_\gamma(t_q) \end{vmatrix}^2 \tag{1.2}$$

where the q summation subscripts $t_1 \dots t_q (1 \leqq q \leqq n)$ independently of each other run through all the ω values of t. From this expression we see that M itself and all its principal minors (of all orders) are non negative. Hence $\mathbf{M}$ is positive definite.

In particular we see that $M=0$ when and only when the determinant

$$\begin{vmatrix} x_1(t_1) \dots x_1(t_n) \\ \dots\dots\dots\dots \\ x_n(t_1) \dots x_n(t_n) \end{vmatrix}$$

vanishes identically in $t_1 \dots t_n$, i. e. when and only when there exists identically in t at least one linear relation

$$a_1x_1(t)+\dots+a_nx_n(t)=0 \tag{1.3}$$

where the coefficients $a_1 \dots a_n$ are independent of t and not all equal to zero. In Sections 6 and 7 the question of linear dependencies in the set $\mathbf{x}$ will be analyzed more closely.

I shall next show that *the determinant value of a positive definite matrix is at most equal to the product of its diagonal elements.*[2]

From this it will follow that

$$0 \leqq M \leqq m_{11} \dots m_{nn}. \tag{1.4}$$

Let $\mathbf{S}$ be a positive definite matrix. If $\mathbf{S}$ is singular, the proposition is trivial. Therefore let $\mathbf{S}$ be non singular, hence all the $s_{ii}>0$. Let us consider one of the real square roots $\mathbf{Q}$ of $\mathbf{S}$, that is one of the real matrices $\mathbf{Q}$ such that $\mathbf{Q}^2=\mathbf{S}$. Since $\mathbf{S}$ is positive definite and non singular, such a matrix always exists, and it can be constructed as a polynomial in $\mathbf{S}$. Being a polynomial in a symmetric argument, $\mathbf{Q}$ is symmetric, therefore $\Sigma_k q_{ik}^2=s_{ii}$.

[1] *Kowalewski:* Determinantentheorie, Leipzig 1909, p. 321.

[2] Comptes rendus de l'Académie des Sciences, Paris. Séance du 5 décembre 1927.

Applying *Hadamard*'s theorem of determinants to the real determinant $Q=|\mathbf{Q}|$, we get $Q^2 \overline{\overline{<}} s_{11} \dots s_{nn}$. But $Q^2=S$, hence $S \overline{\overline{<}} s_{11} \dots s_{nn}$.

Now let us consider the effect on $\mathbf{M}$ of a homogeneous linear transformation performed on the observational variables. *If the observational variables are subject to a homogeneous linear transformation, the moment matrix* $\mathbf{M}$ *is transformed as the matrix of the quadratic form* $\mathbf{xMx}$ *where the symbolic variables* $\mathbf{x}$ *are contragredient to the observational variables.*

In fact, suppose that the observational set $\mathbf{x}$ is subject to the transformation $\mathbf{x}'=\mathbf{Cx}$. The mean values of the new observational variables $\mathbf{x}'$ will evidently be zero (since the mean values of the original variables $\mathbf{x}$ are zero). Let $\mathbf{M}'$ be the moment matrix of the new observational set $\mathbf{x}'$. We have $x_i'=\Sigma_k c_{ik}x_k$, hence

$$(\mathbf{M}')_{ij}=m_{ij}'=\Sigma_i x_i' x_j'=\Sigma_{kh}\Sigma_i c_{ik}x_k x_h \dot{c}_{hj}=\Sigma_{kh}c_{ik}m_{kh}\dot{c}_{hj}.$$

This shows that if the observational set $\mathbf{x}$ with the moment matrix $\mathbf{M}$, is subject to the transformation $\mathbf{x}'=\mathbf{Cx}$, the moment matrix $\mathbf{M}'$ of the new observational set $\mathbf{x}'$ will be

$$\mathbf{M}'=\mathbf{C}\mathbf{M}\dot{\mathbf{C}}. \tag{1.5}$$

But this is just the matrix of the quadratic form $\mathbf{x}'\mathbf{M}'\mathbf{x}'$ into which $\mathbf{xMx}$ is carried when the symbolic variables $\mathbf{x}$ are subject to the transformation $\mathbf{x}=\dot{\mathbf{C}}'\mathbf{x}'$ which is the transformation contragredient to the transformation performed on the observational variables.

Instead of studying linear transformations on the observational variables we can therefore simply study linear transformations of the positive definite quadratic form $\mathbf{xMx}$. This remark will prove to be very useful.

2. THE CORRELATION MATRIX AND THE COEFFICIENT OF COLLECTIVE SCATTER.

If the set $\mathbf{x}$ of observational variables is an effective set, we may consider the normalized of $\mathbf{M}$. This matrix

$$\mathbf{R}=\mathbf{D}^{-\frac{1}{2}}\mathbf{M}\mathbf{D}^{-\frac{1}{2}} \tag{2.1}$$

where $\mathbf{D}$ is the diagonal matrix of $\mathbf{M}$, has as its elements the simple correlation coefficients

$$r_{ij}=+\frac{m_{ij}}{\sqrt{m_{ii}m_{ij}}} \qquad r_{ii}=1. \tag{2.2}$$

The matrix

$$\mathbf{R}=(r_{ij})=\begin{pmatrix} r_{11}\dots r_{1n} \\ \dots\dots\dots \\ r_{n1}\dots r_{nn}\end{pmatrix}$$

will be called the *correlation matrix* of the set $\mathbf{x}$. The determinant value of $\mathbf{R}$ is equal to

$$R=\frac{M}{m_{11}\dots m_{nn}}. \tag{2.3}$$

The positive square root of R

$$s=+\sqrt{R} \tag{2.4}$$

will be called the *collective scatter coefficient* for the set $\mathbf{x}$, or shorter the collective scatter in $\mathbf{x}$. To break monotony s might also be called the coefficient of scatter. The quantity

$$r=+\sqrt{1-R} \tag{2.5}$$

will be called the *collective correlation coefficient* for the set $\mathbf{x}$, or shorter the collective correlation in $\mathbf{x}$. For $n=2$ the collective correlation coefficient reduces to the simple correlation coefficient, if the appropriate sign is attributed to the square root.

The collective scatter and the collective correlation are related by the formula

$$r^2+s^2=1. \tag{2.6}$$

If the simple correlation coefficients were not sometimes called total correlation coefficients, it would have been more natural to call r the total correlation coefficient and s the total scatter coefficient. As it is, it will probably be safer to introduce a new adjective like collective. The significance of s as a scatter coefficient can be visualized in the following way.

Let $\mathbf{z}=(z_1\dots z_n)$ be a point in n dimensional space. If there is given $n+1$ points $\mathbf{z}(t_0)$, $\mathbf{z}(t_1)\dots\mathbf{z}(t_n)$, these points may be taken as defining a corner, the $n+1$ vertices of which are the points $\mathbf{z}(t_i)$. For $n=2$ the corner is a triangle, for $n=3$ a tetraedron.

The volume of the corner defined by these $n+1$ points is equal to

$$V=\frac{1}{n!}\begin{vmatrix} 1 & 1 & \dots & 1 \\ z_1(t_0) & z_1(t_1) & \dots & z_1(t_n) \\ z_2(t_0) & z_2(t_1) & \dots & z_2(t_n) \\ \dots & \dots & \dots & \dots \\ z_n(t_0) & z_n(t_1) & \dots & z_n(t_n) \end{vmatrix}. \tag{2.7}$$

If we introduce the coordinates measured from $\mathbf{z}(t_0)$, that is if we put $\mathbf{x}(t_i)=\mathbf{z}(t_i)-\mathbf{z}(t_0)$, the formula (2.7) reduces to

$$V=\frac{1}{n!}\begin{vmatrix} x_1(t_1) & \dots & x_1(t_n) \\ \dots & \dots & \dots \\ x_n(t_1) & \dots & x_n(t_n) \end{vmatrix}. \tag{2.8}$$

Now suppose we have given a scatter diagram, that is ω points $\mathbf{x}(t)$, where t runs through $t=1,2\dots\omega$. The components of $\mathbf{x}(t)$ are the n attributs of the t-th observation, each attribut being measured from its mean. If we pick out any set of n observation points $\mathbf{x}(t_1)\dots\mathbf{x}(t_n)$ and construct the corner on these n points *and the mean of all observations* as an $(n+1)$-th point, then the volume of this corner is given by (2.8). Fig. 1 shows 3 such corners constructed in an $n=2$ dimensional scatter diagram.

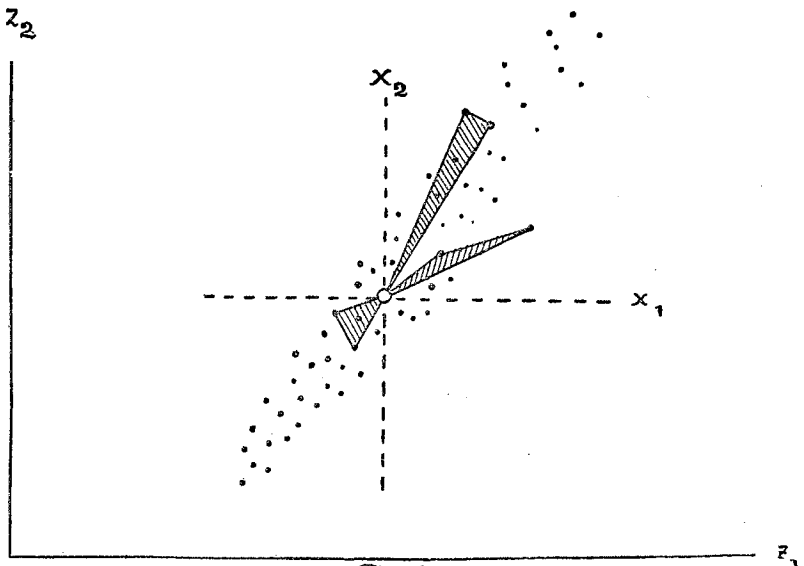

Fig. 1.

If there exists identically in t a relation of the form (1.3), that is if the ω observation points are rigorously lying in a plane in

the n dimensional scatter diagram, the volume (2.8) will be zero *for any set of n observation points* we might choose. If the ω observation points are not lying rigorously in a plane, the volume (2.8) will be different from zero for some of the sets (perhaps for all of them). The magnitude of the volume (2.8) for any set of n points, offers a measure of how far the corner constructed on these n points, is from being flattened down to the shape of a plane.

The idea therefore naturally presents itself to adopt the *average value of all these volumes*, taken over all possible sets of n points, as a measure of how far the swarm of observation points is from lying in a plane. The formula (2.8) gives a volume which may be either positive or negative, the sign of the volume being defined by a convention as to the sequence of the axes. In order not to have positive and negative volumes cancelled out, we take the square mean σ defined by

$$\sigma^2=\frac{1}{n!\,\omega^n}\Sigma_{t_1\dots t_n}\begin{vmatrix} x_1(t_1) & \dots & x_1(t_n) \\ \dots & \dots & \dots \\ x_n(t_1) & \dots & x_n(t_n) \end{vmatrix}^2 \tag{2.9}$$

where the n summation subscripts $t_1\dots t_n$ independently of each other run through all the ω values of t and the square root is taken positive. The quantity σ defined by (2.9) might be called the *collective standard deviation* in the set $\mathbf{x}$. For $n=1$, σ reduces to the ordinary standard deviation on a single variable.

The coefficient σ is an *absolute* quantity in the sense that it depends on the units with which the variables are meaured. More precisely: if any of the variables is multiplied by a constant, σ is multiplied by that same constant. A *relative* measure of the lack of linearity in the set $\mathbf{x}$ will therefore be obtained if σ is divided by the product of the standard deviations of the individual variables. This ratio $s=\dfrac{\sigma}{\sigma_1\dots\sigma_n}$ is nothing else than the scatter coefficient s defined by (2.4). This simply follows from (1.2) and (2.3).

In this and the following sections we shall establish other properties of s, which makes it further plausible to adopt the closeness of s to zero as a measure of how close the set $\mathbf{x}$ comes to being linearly dependent.

The adjoint $\hat{\mathbf{R}}$ of $\mathbf{R}$ will be called the *adjoint correlation matrix*, and the positive square root $+\sqrt{\hat{r}_{ii}}$ of the i-th diagonal element

in $\hat{\mathbf{R}}$ will be called the *coefficient of linear importance* of x_i in the set $\mathbf{x}$.

Taking the adjoint of (2.1) we get

$$\hat{\mathbf{R}} = \frac{\mathbf{D}^{\frac{1}{2}}\hat{\mathbf{M}}\mathbf{D}^{\frac{1}{2}}}{m_{11}\dots m_{nn}} \tag{2.10}$$

that is

$$\hat{r}_{ij} = \frac{\sqrt{m_{ii}m_{jj}}}{m_{11}\dots m_{nn}}\hat{m}_{ij} \tag{2.11}$$

where the square root is taken positive.

The q rowed principal submatrix $\mathbf{R}_{q[\alpha\dots\gamma]}$ formed by the q rows $\alpha\dots\gamma$ and the q columns $\alpha\dots\gamma$ from $\mathbf{R}$, is the correlation matrix, and the positive square root of its determinant value is the scatter coefficient for the subset $x_\alpha\dots x_\gamma$. This determinant value is equal to

$$R_{q[\alpha\dots\gamma]} = \frac{M_{q[\alpha\dots\gamma]}}{m_{\alpha\alpha}\dots m_{\gamma\gamma}}. \tag{2.12}$$

The matrix $\mathbf{N}=\omega\mathbf{R}$ is the moment matrix of the standardized set $\mathbf{y}$. Since $\omega\mathbf{R}$ is the moment matrix for a set of statistical variables, $\omega\mathbf{R}$ and hence $\mathbf{R}$ is positive definite, and being positive definite, it has a determinant value at most equal to the product of its diagonal elements. Therefore

$$0 \leqq R \leqq 1 \text{ and } 0 \leqq s \leqq 1. \tag{2.13}$$

Since $R_{q[\alpha\dots\gamma]}$ has the same significance for the set $x_\alpha\dots x_\gamma$ as R has for the set $x_1\dots x_n$, it is seen that any principal minor in R has a value between 0 and 1.

We even have the following proposition: *Any principal minor in R is greater than or at least equal to R. More generally: If R' is a principal minor in R, then any principal minor contained in R' is greater than or at least equal to R'.*

It is understood that any of the principal minors considered shall be formed by picking out q *different* rows (and the corresponding q columns) from R. If we would consider a principal minor formed by picking out any row (and the corresponding column) of R, twice or several times, we should have to compare this minor with the determinant obtained from R by counting the row (and column) in question just as many times as it is

counted in the minor. This is, however, a trivial case, since in this case both the minor and the determinant in question would be zero.

Our proposition will be proved if we can prove the lemma that any q rowed principal minor R_q in R is greater than or at least equal to the $(q+1)$ rowed principal minor obtained by adding an arbitrary row (and the corresponding column) to R_q.

For the sake of brevity let the principal minor which is formed by the q rows $\alpha\dots\gamma$ and the q columns $\alpha\dots\gamma$ from R, be detoned $[\alpha\dots\gamma]$. From a fundamental theorem regarding the relations between the minors in a matrix we have

$$[\alpha\dots\gamma i]\cdot[\alpha\dots\gamma j]-[\alpha\dots\gamma]\cdot[\alpha\dots\gamma ij]=\triangle_{ij}\triangle_{ji}$$

where $\triangle_{ij}$ is the determinant obtained by adding the i-th row and the j-th column to $[\alpha\dots\gamma]$. Since $\mathbf{R}$ is symmetric and real, $\triangle_{ij}\triangle_{ji}=\triangle_{ij}^2\geqq 0$. We therefore have

$$[\alpha\dots\gamma i]\cdot[\alpha\dots\gamma j]\geqq[\alpha\dots\gamma]\cdot[\alpha\dots\gamma ij]. \tag{2.14}$$

Now our lemma is evidently true for $q=1$ (because $r_{ii}=1$ is not less than $1-r_{ij}^2$), and from (2.14) it follows that if the lemma is true for q, it is also true for $q+1$. In fact if it is true for q, we have $[\alpha\dots\gamma]\geqq[\alpha\dots\gamma j]$ $(j=1,2\dots n)$, and hence from (2.14) if $[\alpha\dots\gamma]$ is different from zero: $[\alpha\dots\gamma i]\geqq[\alpha\dots\gamma ij]$, $(i,j=1,2\dots n)$. The last relation also holds good if $[\alpha\dots\gamma]=0$. For in this case $[\alpha\dots\gamma i]$ and $[\alpha\dots\gamma ij]$, being positive definite determinants with a vanishing principal minor, must both be zero.

Therefore any $(q+1)$ rowed principal minor is greater than or at least equal to the $(q+2)$ rowed principal minor obtained by adding an arbitrary row (and the corresponding column) to it. This establishes our proposition.

The last proposition can also be stated thus: *The coefficient of scatter for a set of variables is never greater than the coefficient of scatter for a subset contained in the set.* The more variables we take into account, the greater is therefore the possibility that the variables shall come close to being linearly dependent, the sense of "coming close to" being defined by the magnitude of the scatter coefficient.

An interesting application of the preceding proposition is to compare R with its two rowed principal minors. This gives

$$1-r_{ij}^2\geqq R \qquad (i\neq j).$$

If r_{ij} is any of the simple correlation coefficients in the set $\mathbf{x}$, we therefore always have

$$|r_{ij}| \leqq +\sqrt{1-R} = r \qquad (i \neq j). \tag{2.15}$$

That is, the collective correlation in the set $\mathbf{x}$ is never less than the absolute value of any of the simple correlations in the set.

We might also consider the collective correlation in any subset $x_\alpha \ldots x_\gamma$. And we would have the proposition that *the collective correlation in a set is never less than the collective correlation in a subset contained in the set.*

The discussion of the following sections is formulated in terms of the scatter coefficient, but it would have been equally possible to formulate it in terms of the collective correlation coefficient.

3. UNCORRELATED VARIABLES AND ORTHOGONAL TRANSFORMATIONS.

The n variables $x_1 \ldots x_n$ are called *uncorrelated* or *orthogonal* when all the product moments m_{ij} (taken about the means) are equal to zero for $i \neq j$, i. e. when the moment matrix $\mathbf{M}$ is a diagonal matrix. This formulation of the definition makes it applicable also to the case where some of the variables are identically zero.[1]

If none of the variables are identically zero, i. e. if all the variables are effective, the preceding definition is equivalent with the following: The n variables $x_1 \ldots x_n$ are uncorrelated when all the simple correlation coefficients r_{ij} $(i \neq j)$ are equal to zero, i. e. when the correlation matrix $\mathbf{R}$ is equal to the unit matrix $\mathbf{E}$.

The n effective variables $x_1 \ldots x_n$ are uncorrelated when and only when the correlation matrix is orthogonal. In other words: $\mathbf{R}$ is orthogonal when and only when $\mathbf{R} = \mathbf{E}$. In fact if $\mathbf{R} = \mathbf{E}$, $\mathbf{R}$ is evidently orthogonal. Inversely if $\mathbf{R}\tilde{\mathbf{R}} = \mathbf{E}$, we must have $\Sigma_j r_{ij}^2 = 1$, which is only possible when $r_{ij} = 0$ $(i \neq j)$, hence $\mathbf{R} = \mathbf{E}$.

The n effective variables $x_1 \ldots x_n$ are uncorrelated when and only when the coefficient of scatter s is equal to 1. For if $\mathbf{R} = \mathbf{E}$, we evidently have $R = 1$. Inversely if $R = 1$, we must have $r_{ij} = 0$ $(i \neq j)$ by virtue of (2. 15). Hence $\mathbf{R} = \mathbf{E}$.

[1] The notion of non correlation should be distinguished from the notion of independence. The n variables $x_1 \ldots, x_n$ may be called independent if the frequency function $P(x_1 \ldots x_n)$ is of the form $P(x_1 \ldots x_n) = P_1(x_1) \ldots P_n(x_n)$ where the P_i are functions of a single variable. Independent variables are always uncorrelated, but the inverse is not true.

From the two preceding propositions follows:

The correlation matrix $\mathbf{R}$ is orthogonal when and only when the coefficient of scatter s is equal to 1.

We shall further establish a property of the coefficient of scatter which relates this coefficient to the mutual inclination of the normals on the n regressionplanes. We shall make use of the following propositions which will be established in Sections 4 and 6: There exist n distinct regression planes when and only when $R \neq 0$. If the n regression planes exist, the normal on the i-th regression plane is a vector whose components are proportional to the elements in the i-th row of the adjoint moment matrix $\hat{\mathbf{M}}$ (which is now non singular). Using this we have the proposition:

The n regression normals are mutually orthogonal when and only when the coefficient of scatter s is equal to 1, i. e. when and only when the n variables $x_1 \ldots x_n$ are uncorrelated.

Since the i-th regression normal is a vector whose components are proportional to the elements in the i-th row of $\hat{\mathbf{M}}$, the regression normals are mutually orthogonal when and only when $\hat{\mathbf{M}}\check{\mathbf{M}} = \hat{\mathbf{M}}^2$ is a diagonal matrix. In this case $\hat{\mathbf{M}}$ must also be a diagonal matrix. In order to prove this, we shall establish the proposition:

The square of a positive definite and non singular matrix $\mathbf{S}$ is a diagonal matrix when and only when $\mathbf{S}$ itself is a diagonal matrix.

If $\mathbf{S}$ is diagonal, $\mathbf{S}^2$ is evidently also diagonal. Inversely suppose that $\mathbf{S}^2$ is a diagonal matrix. Since $\mathbf{S}$ is positive definite, we have $S \leqq s_{11} \ldots s_{nn}$. Further the determinant value of the diagonal matrix $\mathbf{S}^2$ is equal to the product of its diagonal elements. Hence we must have

$$S^2 = |\mathbf{S}^2| = \Sigma_k s_{1k}^2 \ldots \Sigma_k s_{nk}^2 \leqq s_{11}^2 \ldots s_{nn}^2$$

where none of the s_{ii} are zero (since $\mathbf{S}$ is non singular). But this is only possible if $s_{ij} = 0$ $(i \neq j)$. Therefore $\mathbf{S}$ must be a diagonal matrix.

The n regression normals are therefore mutually orthogonal when and only when $\hat{\mathbf{M}}$ is diagonal, that is when and only when $\mathbf{M}$ is diagonal, hence when and only when $\mathbf{R} = \mathbf{E}$.

In some cases it may be of interest to reduce a set of statistical variables to an uncorrelated form by means of a non singular linear transformation. If this is done, the correlation between the actually observed variables may be looked upon as due to the

fact that each of the observed variables is a linear combination of certain underlying elementary variables, that are uncorrelated[1].

Such a reduction is always possible and in an infinity of ways. In fact the moment matrix $\mathbf{M}$ for a set of observational variables is transformed as the matrix of the quadratic form $\mathbf{xMx}$, where the symbolic variables $\mathbf{x}$ are contragredient to the observational variables. The problem is therefore only to determine a real and non singular matrix $\mathbf{C}$ such that $\mathbf{CM\grave{C}}$ is a diagonal matrix. If $\mathbf{C}$ has this property, the transformation $\mathbf{x}' = \mathbf{Cx}$ performed on the given observational set $\mathbf{x}$ will yield a set of uncorrelated variables $\mathbf{x}'$.

Now, if $R \neq 0$ (hence $\mathbf{M}$ non singular), the existence of a real and non singular $\mathbf{C}$ such that $\mathbf{CM\grave{C}} = \mathbf{E}$, simply follows from the fact that $\mathbf{M}$, being positive definite and non singular, has a real square root $\mathbf{M}^{\frac{1}{2}}$, which is also symmetric and non singular. This being so, $\mathbf{C} = \mathbf{M}^{-\frac{1}{2}}$ is such that $\mathbf{CM\grave{C}} = \mathbf{E}$, for we have $\mathbf{CM\grave{C}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}} = \mathbf{E}$.

More generally, the problem here considered, is simply the problem of reducing the real quadratic form $\mathbf{xMx}$ to a sum of squares by means of a real and non singular linear transformation. And from the theory of quadratic forms we know that this is always possible. It is even possible to find a real orthogonal (and hence non singular) matrix $\mathbf{O}$ such that $\mathbf{OM\grave{O}}$ is a diagonal matrix. The elements of this diagonal matrix will be the roots of the secular equation for $\mathbf{M}$. These are all non negative since $\mathbf{M}$ is positive definite. The number of these roots which are equal to zero, is $n - \varrho$, where ϱ is the rank of $\mathbf{M}$. Hence $n - \varrho$ of the variables in the uncorrelated set must have a standard deviation equal to zero, and therefore be identically zero (since they are measured from their means). The same holds good generally, not only for orthogonal transformations. For let $\mathbf{M}$ be of rank ϱ and $\mathbf{C}$ a non singular matrix. The product $\mathbf{CM\grave{C}}$ is also of rank ϱ. If it is a diagonal matrix, $n - \varrho$ of its diagonal elements must therefore be zero. Hence we have the proposition:

If a set of statistical variables is reduced to an uncorrelated set by means of a non singular linear transformation, then this uncorrelated set will contain exactly ϱ effective variables (i. e. ϱ variables with a standard deviation different from zero), where ϱ is the rank of the moment matrix $\mathbf{M}$ of the original set.

Now suppose we have determined a non singular transformation

[1] This point of view was suggested to me by Professor *T. L. Kelley* of Stanford University.

$\mathbf{x}' = \mathbf{Cx}$ which carries the given observational set $\mathbf{x}$ over into the uncorrelated set $\mathbf{x}'$. Then $\mathbf{CM\grave{C}}$ is a diagonal matrix $\begin{pmatrix} d_1 \dots 0 \\ \dots\dots \\ 0 \dots d_n \end{pmatrix}$. Let $d_\alpha \dots d_\gamma$ be the $n - \varrho$ diagonal elements in this matrix, which are equal to zero. If in the expression $\mathbf{x} = \mathbf{C}^{-1}\mathbf{x}'$ we put the $n - \varrho$ variables $x'_\alpha \dots x'_\gamma$ identically equal to zero, we obtain a ϱ-parametric expression for the observational variables $\mathbf{x}$, which holds good identically in t, the ϱ parameters being uncorrelated. The set of these ϱ parameters, that is the set obtained from the set $\mathbf{x}'$ by omitting $x'_\alpha \dots x'_\gamma$ might be called a *parameter* set for the set $\mathbf{x}$. If we further standardize the ϱ variables in the parameter set, we obtain a set which might be called an *elementary* set for the set $\mathbf{x}$. Thus an elementary set for $\mathbf{x}$ is a set obtained by omitting the ineffective and standardizing the effective variables in any uncorrelated set obtained from $\mathbf{x}$ by a non singular linear transformation.

There exists an infinity of elementary sets for any given set $\mathbf{x}$: *Any ϱ dimensional set obtained from an elementary (and therefore ϱ dimensional) set by means of an arbitrary orthogonal transformation (in ϱ variables) is also an elementary set.*

In order to prove this let us consider the n dimensional set $\mathbf{y}' = (y'_1 \dots y'_n)$ obtained from the observational set $\mathbf{x}'$ by leaving the variables $x'_\alpha \dots x'_\gamma$ unchanged, and standardizing the other ϱ variables in the set. The set $\mathbf{y}'$ can be looked upon as being obtained from $\mathbf{x}'$ and hence from $\mathbf{x}$ by a non singular linear transformation in n variables.

The n rowed moment matrix $\mathbf{N}'$ for the set $\mathbf{y}'$ is $\mathbf{N}' = \omega\,\mathbf{E}'$, where $\mathbf{E}'$ is the n rowed matrix obtained from the unit matrix $\mathbf{E}$ by replacing the $n - \varrho$ diagonal elements nos. $\alpha \dots \gamma$ by zeros.

Let us consider an n rowed matrix of the following type: The elements in the $n - \varrho$ rows $\alpha \dots \gamma$ and the $n - \varrho$ columns $\alpha \dots \gamma$ shall consist exclusively of zeros, with exception of the diagonal elements in these rows and columns; these diagonal elements are put equal to 1. The elements of the remaining ϱ rows and ϱ columns, taken by themselves, shall form an arbitrary ϱ rowed orthogonal matrix. The n rowed matrix thus formed is also an orthogonal (and hence non singular) matrix $\mathbf{O}$. Now let us perform on the observational set $\mathbf{y}'$ the transformation $\mathbf{y}'' = \mathbf{Oy}'$. By this transformation we have $y''_\alpha = y'_\alpha \dots y''_\gamma = y'_\gamma$ and the remaining ϱ variables, taken by themselves, are transformed orthogonally. Furthermore, the n dimensional set $\mathbf{y}''$ can be looked

upon as being obtained from the given set $\mathbf{x}$ by a non singular transformation. The moment matrix $\mathbf{N}''$ for the set $\mathbf{y}''$ is $\mathbf{N}'' = \mathbf{O}\mathbf{N}'\dot{\mathbf{O}} = \omega\mathbf{O}\mathbf{E}'\dot{\mathbf{O}}$. Now it is evident from the way $\mathbf{O}$ and $\mathbf{E}'$ are formed, that $\mathbf{O}\mathbf{E}'\dot{\mathbf{O}} = \mathbf{E}'$, hence $\mathbf{N}'' = \mathbf{N}'$. The ϱ dimensional set which remains when $y''_\alpha \dots y''_\gamma$ are omitted from the set $\mathbf{y}''$, is therefore both an uncorrelated and a standardized ϱ dimensional set. Hence it is an elementary set for $\mathbf{x}$. This establishes the proposition.

One method of actually performing the reduction of a given set to an uncorrelated form, is the following, which is applicable whenever the coefficient of scatter s is different from zero.

Let us consider the sequence of matrices $\mathbf{R}^{(\nu)} = (r^{(\nu)}_{ij})$ $(i,j = 1, 2 \dots \nu;$ $\nu = 1, 2 \dots n)$, defined by the recurrence formula

$$r^{(\nu-1)}_{ij} = r^{(\nu)}_{\nu\nu} \cdot r^{(\nu)}_{ij} - r^{(\nu)}_{i\nu} \cdot r^{(\nu)}_{\nu j}$$
$$(\nu = n, n-1, \dots 2;\ i,j = 1, 2 \dots \nu)$$
$$r^{(n)}_{ij} = r_{ij}$$

where r_{ij} are the simple correlation coefficients in the given set $\mathbf{x}$. The matrix $\mathbf{R}^{(\nu)}$ is ν rowed and symmetric.

Now put $u_{ij} = r^{(j)}_{ij}$ and consider the n rowed matrix $\mathbf{U} = (u_{ij})$. All the elements of $\mathbf{U}$ below the principal diagonal are zero. If $R \neq 0$, the matrix $\mathbf{U}$ solves the problem of reducing the set $\mathbf{x}$ to an uncorrelated form. We have the proposition:

If the coefficient of scatter for a set of statistical variables is different from zero, the matrix $\mathbf{U}$, *defined above, is non singular, and all the elements of its principal diagonal are positive, not zero. Let the given set be reduced to standard coordinates* $\mathbf{y}$. *If the set* $\mathbf{y}$ *is expressed linearly in terms of the set* $\mathbf{y}'$ *by the relations*

$$\mathbf{y} = \mathbf{U}\mathbf{y}'$$

then the observational set $\mathbf{y}'$ *will be an uncorrelated set, and the standard deviation of* y'_i *will be equal to*

$$\frac{1}{\sqrt{u_{ii} \cdot u_{i+1,\,i+1} \dots u_{nn}}}.$$

In order to prove this we shall consider the sequence of quadratic forms

$$f_\nu = \Sigma_{ij}\, y_i\, r^{(\nu)}_{ij}\, y_j$$
$$(\nu = 1, 2 \dots n)$$

f_ν is a quadratic form involving only the ν variables $y_1 \dots y_\nu$. The variables are now symbolic, not observational.

Let us put

$$y'_\nu = \Sigma_j\, r^{(\nu)}_{\nu j}\, y_j$$
$$(\nu = 1, 2 \dots n)$$

This is equivalent with

$$\mathbf{y}' = \dot{\mathbf{U}}\mathbf{y}$$

y'_ν is a linear form involving only the ν variables $y_1 \dots y_\nu$.

By introducing the expression for y'_ν it is easily seen that the equation

$$u_{\nu\nu} f_\nu = y'^2_\nu + f_{\nu-1} \tag{3.1}$$

holds good identically in all the variables involved.

The equation (3.1) shows that if f_ν is a positive definite and non singular form (hence $u_{\nu\nu} > 0$), $f_{\nu-1}$ must be the same. For suppose $f_{\nu-1}$ were not. By a real and non singular transformation performed on its $\nu-1$ variables $y_1 \dots y_{\nu-1}$ we could write $f_{\nu-1}$ as a sum of $\nu-1$ squares, and at least one of these squares would have a non positive coefficient. Hence f_ν would be written as a sum of ν squares, at least one coefficient of which is non positive. Furthermore, this expression for f_ν can be looked upon as being obtained by a real and non singular transformation performed on its ν variables $y_1 \dots y_\nu$. In fact the expression considered would be obtained by a transformation whose ν rowed matrix is of the following form: The last element of its principal diagonal is $u_{\nu\nu} > 0$. The rest of the elements in the ν-th column are zero, and the rest of the elements in the ν-th row are the quantities $r^{(\nu)}_{\nu j}$. Further the elements of the first $\nu-1$ rows and columns, taken by themselves, is the matrix of the real and non singular transformation performed on the $\nu-1$ variables in $f_{\nu-1}$. The determinant value of this $\nu-1$ rowed matrix is different from zero. Hence the transformation performed on f_ν is real and non singular. This shows that $f_{\nu-1}$ must be positive definite and non singular when f_ν has these properties.

Since in our case (i. e. $R \neq 0$) the quadratic form $f_n = \mathbf{y}\mathbf{R}\mathbf{y}$ is positive definite and non singular, it follows that all the forms f_ν are positive definite and non singular. Hence all the principal minors of all the matrices $\mathbf{R}^{(\nu)}$ are positive, not zero. In particular all the elements of their principal diagonals, and therefore

all the quantities $u_{11}, u_{22}, \ldots u_{nn}$ are positive, not zero. Since the determinant $|\mathbf{U}| = u_{11} \ldots u_{nn}$ is positive, we see that $\mathbf{U}$ is non singular.

From equation (3.1) follows

$$f_n = \mathbf{yRy} = \sum_{i=1}^{n} \frac{y_i'^2}{u_{ii} u_{i+1,\, i+1} \ldots u_{nn}}.$$

Introducing the matrix $\mathbf{N} = \omega \mathbf{R}$, we can state the result thus: The transformation $\mathbf{y}' = \grave{\mathbf{U}}\mathbf{y}$ performed on the symbolic variables in the quadratic form $\mathbf{yNy}$ carries this form over into a sum of squares, the i-th coefficient in the sum being equal to

$$\frac{\omega}{u_{ii} \ldots u_{nn}}. \tag{3.2}$$

In other words $\mathbf{U}^{-1}\mathbf{N}\grave{\mathbf{U}}^{-1}$ is a diagonal matrix whose i-th diagonal element is equal to (3.2). Now $\mathbf{N}$ is the moment matrix of the observational set $\mathbf{y}$. On account of the contragredience between the symbolic and the observational variables we therefore have: If the transformation $\mathbf{y} = \mathbf{Uy}'$ is performed on the standardized observational set $\mathbf{y}$, whose moment matrix is $\mathbf{N}$, we obtain an observational set $\mathbf{y}'$ whose moment matrix $\mathbf{N}' = \mathbf{U}^{-1}\mathbf{N}\grave{\mathbf{U}}^{-1}$ is a diagonal matrix, the i-th diagonal element of which is (3.2). This establishes our proposition.

From the possibility of reducing any observational set $\mathbf{x}$ to an uncorrelated form by means of a real and non singular transformation, follows that if $R \neq 0$, it is always possible to increase R by a real and non singular transformation. It is even possible to make R rigorously equal to 1.

Inversely: If $R \neq 0$, it is always possible to lower R by a real and non singular transformation. It is even possible to bring R as near to zero as we please. But it is not possible by a non singular transformation to make R rigorously equal to 0. In fact, if the transformation $\mathbf{x}' = \mathbf{Cx}$ is performed on the observational set $\mathbf{x}$, the coefficient of scatter for the set $\mathbf{x}'$, will be given by

$$R' = \frac{m_{11} \ldots m_{nn}}{m_{11}' \ldots m_{nn}'} C^2 R$$

where m_{ii} and m_{ii}' are the diagonal elements in the moment matrices $\mathbf{M}$ and $\mathbf{M}'$ of $\mathbf{x}$ and $\mathbf{x}'$ respectively, and $C = |\mathbf{C}|$ is the modulus of transformation. If $\mathbf{C}$ is non singular, none of the factors in the expression for R' is zero. Hence R' is different from zero. But it is possible to bring R' as near to zero as we please. For there exists an infinity of singular transformations which reduce R' to zero. In fact, under the transformation $\mathbf{x}' = \mathbf{Cx}$, we have $m_{ii}' = \mathbf{c}_i \mathbf{M} \mathbf{c}_i$, where $\mathbf{c}_i$ is the i-th direct (vector) component of $\mathbf{C}$. Since m_{ii}' is a positive definite and non singular quadratic form in the n variables $c_{i1} \ldots c_{in}$, m_{ii}' is always positive, not zero, except for $c_{i1} = \ldots = c_{in} = 0$. A necessary and sufficient condition that none of the diagonal elements of $\mathbf{M}'$ shall be zero, is therefore that none of the rows of $\mathbf{C}$ shall consist exclusively of zeros. And it is evidently possible to construct a singular $\mathbf{C}$, where none of the rows consist exclusively of zeros. Any such $\mathbf{C}$ would reduce R' to 0.

4. THE REGRESSION EQUATIONS.

The regression equations are usually defined as the n homogeneous linear equations

$$\sum_j b_{ij} x_j = 0 \qquad b_{ii} = -1 \qquad (i = 1,2 \ldots n) \tag{4.1}$$

which are obtained by expressing in turn each of the variables x_i in terms of the other variables, and determining the constant coefficients b_{ij} $(i \neq j)$ so as to minimize the sum square of the deviations

$$\sum_t [\sum_j b_{ij} x_j(t)]^2 \qquad (i = 1,2 \ldots n).$$

Introducing the matrix

$$\mathbf{B} = (b_{ij}) = \begin{pmatrix} b_{11} \ldots b_{1n} \\ \ldots\ldots\ldots \\ b_{n1} \ldots b_{nn} \end{pmatrix} \qquad \text{where } b_{ii} = -1$$

the regression equations (4.1) may be written

$$\mathbf{Bx} = 0. \tag{4.2}$$

This should not be interpreted as a system of linear equations in $\mathbf{x}$. In fact if $|\mathbf{B}| \neq 0$, the system would have no solution, except $\mathbf{x} = 0$; (4.2) is only a symbolic way of expressing the n regression equations.

The coefficients b_{ij} are restricted by the conditions $b_{ii}=-1$. If a general analysis is to be made, it is desirable to adopt a slightly different definition of the regression equations so as to remove the condition $b_{ii}=-1$. This is done by writing the regression equations in the form

$$\Sigma_j a_{ij}x_j=0 \qquad (i=1,2\ldots n) \tag{4.3}$$

where the a_{ii} are constants later to be disposed of (not to be determined by the minimizing conditions). The difference between the regression systems (4.1) and (4.3) is not only formal. It might happen that the minimizing conditions satisfied by the a_{ij} $(i\neq j)$, makes it impossible to put $a_{ii}=-1$. But if this is possible, the systems (4.1) and (4.3) will be equivalent.

Introducing the matrix

$$\mathbf{A}=(a_{ij})=\begin{pmatrix} a_{11}\ldots a_{1n} \\ \ldots\ldots\ldots \\ a_{n1}\ldots a_{nn}\end{pmatrix}$$

the regression equations (4.3) may be written

$$\mathbf{Ax}=0. \tag{4.4}$$

When I use the expressions regression matrix and regression equations, without further specification, I always refer to the matrix $\mathbf{A}$ and the equations $\mathbf{Ax}=0$. The matrix $\mathbf{B}$ and the equations $\mathbf{Bx}=0$ will be referred to as the reduced regression matrix and the reduced regression equations.

If $\mathbf{A}$ is known and all the $a_{ii}\neq 0$, we can evidently pass back from $\mathbf{A}$ to $\mathbf{B}$ by the formula

$$b_{ij}=-a_{ij}/a_{ii}. \tag{4.5}$$

The problem of the regression equations is to determine the matrix $\mathbf{A}$. This problem is very easily solved by the elementary rules of matrix algebra.

In fact, the necessary minimizing conditions are

$$\frac{\partial \Sigma_t[\Sigma_k a_{ik}x_k(t)]^2}{\partial a_{ij}}=0 \qquad (i\neq j)$$

that is

$$\Sigma_k a_{ik}m_{kj}=0 \qquad (i\neq j).$$

Therefore if we dispose of the constants a_{ii} in such a way that

$$\Sigma_k a_{ik}m_{ki}=M \qquad (i=1,2\ldots n)$$

the minimizing conditions will be

$$\mathbf{AM}=M\mathbf{E}.$$

If $\mathbf{M}$ is non singular, we can solve for $\mathbf{A}$ by postmultiplication with $\mathbf{M}^{-1}$, which gives

$$\mathbf{A}=\hat{\mathbf{M}} \tag{4.6}$$

where $\hat{\mathbf{M}}$ is the adjoint moment matrix. If the minimizing problem has a solution, the n regression equations must therefore be

$$\Sigma_j \hat{m}_{ij}x_j=0 \qquad (i=1,2\ldots n). \tag{4.7}$$

That this really furnishes a solution of the minimizing problem, is seen by the following argument. The ν-th regression equation (that is the equation (4.7) for $i=\nu$) is determined by minimizing the function

$$f_\nu=\Sigma_t[\Sigma_k a_{\nu k}x_k(t)]^2=\Sigma_{kh}a_{\nu k}m_{kh}a_{\nu h}$$

where $a_{\nu 1}\ldots(\text{except } a_{\nu\nu})\ldots a_{\nu n}$ are considered as $n-1$ independent variables, and $a_{\nu\nu}$ as a constant. The relation by which we have disposed of the constant $a_{\nu\nu}$, should be looked upon as a relation introduced after the derivations have been performed. The relation in question does therefore not define $a_{\nu\nu}$ as a function of the $n-1$ independent variables. That is, we have $\partial f_\nu/\partial a_{\nu\nu}=0$ and $\partial f_\nu/\partial a_{\nu i}=2\Sigma_k a_{\nu k}m_{ki}$ $(i\neq\nu)$. And consequently

$$\frac{\partial^2 f_\nu}{\partial a_{\nu i}\partial a_{\nu j}}=2m_{ij} \qquad (i\neq\nu, j\neq\nu).$$

A part from the positive factor 2, the $(n-1)$ rowed matrix of the second order partial derivatives of f_ν is therefore simply the matrix obtained from $\mathbf{M}$ by omitting the ν-th row and the ν-th column. This matrix is always positive definite. And it is furthermore non singular if $\mathbf{M}$ is non singular. Each of the regressions (4.7) will therefore always correspond to a *minimum* of the sum square in question. And furthermore correspond to a *proper*

minimum if $\mathbf{M}$ is non singular, which is the assumption under which (4.6) was derived. In the last part of this section we shall see that the formulae (4.6) and (4.7) actually hold good under a more general condition. The n regressions (4.7) might be called the *elementary* regressions in distinction to certain other types of regressions which will be discussed later. In particular the i-th regression in (4.7) might be called the i-th elementary regression.

If $R \neq 0$ (i. e. $\mathbf{M}$ non singular), the n regression planes determined by $\mathbf{A}$ are all different. In fact the normals on the regression planes are the directions of the direct (vector) components of $\hat{\mathbf{M}}$. If two (or more) of these directions should coincide, $\hat{\mathbf{M}}$ and hence $\mathbf{M}$ would be singular.

If $\mathbf{M}$ is non singular, $\mathbf{B}$ exists and is given by (4.5). It should be noticed that the regression matrix $\mathbf{A}$ is symmetric, while the reduced matrix $\mathbf{B}$ is in general not symmetric.

By virtue of (2.10) the regressions $\hat{\mathbf{M}}\mathbf{x} = 0$ can be written $\hat{\mathbf{R}}\mathbf{D}^{-\frac{1}{2}}\mathbf{x} = 0$, where $\hat{\mathbf{R}}$ is the adjoint correlation matrix. Introducing from (1.1), we see that the regressions in standard coordinates $\mathbf{y}$ will be

$$\hat{\mathbf{R}}\mathbf{y} = 0$$

that is

$$\Sigma_j \hat{r}_{ij} y_j = 0 \qquad (i = 1,2 \ldots n). \tag{4.8}$$

Since the i-th elementary regression is the regression obtained by minimizing the deviations from linearity, measured in x_i direction, it can be looked upon as the regression obtained by attributing the actual lack of agreement with the postulated analytic relation, exclusively to the variation of x_i. Apart from any special significance the n regressions (4.7) might have in the case of a normal distribution, the difference between these regressions is therefore essentially a *difference in assumption*, namely in the assumption regarding the nature of the variability of the variables involved.

I would like to emphasize this point because the real significance of the difference between the n elementary regressions is not always kept sufficiently clear, it seems to me, when it comes to practical applications. This is particulary true, I think, regarding applications in the economic field.

When one of the variables, say x_i, for one reason or another,

happens to have attracted a particular attention in the setting of the problem, the investigator frequently has a tendency to think of the regression in question as written with x_i on the left hand side, and the other variables on the right hand side of the equation sign. And having this form of the regression in mind, he often draws the conclusion that the regression which should be chosen, is the regression of x_i on the other variables, that is the i-th regression in the system (4.7). Most of the time there is perhaps not even question of a conscious conclusion. The second step: picking out the i-th elementary regression, follows mechanically as soon as the attention of the investigator for one reason or another, has ben focussed upon the variable x_i.

Such a principle for choosing one particular of the regressions (4.7) is fallacious, it seems to me. The procedure of determining statistically an analytic relation between n variables, and the procedure of rearranging the terms in a statistically determined relation between the variables, are two things which should be kept distinctly separated. *The choice of regression essentially represents a problem by itself, and should not be confused with the choice of a particular form in which to write the regression chosen.* Nor can we limit the problem to concern a choice between the n alternatives in (4.7). In fact, a great number of regression problems is of such a nature that none of the regressions (4.7) can be accepted as plausible. This is particularly true in the economic field. Here the nature of the problem is most frequently such that it is out of the question (or ought to be out of the question) to adopt a regression procedure which treats the variables unsymmetrically to the extent of attributing all deviations from linearity (or from some other analytic relation) to one particular of the variables. What we need in these cases is some kind of a *mean regression plane.*

Without pretending to go into an extensive and systematic analysis of this important problem, I shall venture a few remarks on some points of the theory. In the present section I shall consider two particular forms of mean regressions, namely first the *orthogonal mean regression*, that is the mean regression plane determined by minimizing the sum squares measured perpendicularly to the plane, and second, a certain type of mean regression which might be called the *diagonal mean* regression. In the next section, adopting a more general point of view, I shall consider the problem of invariance of regression planes, and con-

struct two regressions which are invariant for a general linear homogeneous transformation.

The equation of a mean regression plane in its general form can be written

$$g_0 + g_1 z_1 + \ldots + g_n z_n = 0$$

or shorter

$$g_0 + \mathbf{g}\mathbf{z} = 0 \tag{4.9}$$

where $\mathbf{g} = (g_1 \ldots g_n)$. The coefficients $g_0 g_1 \ldots g_n$ are constants independent of t; $\mathbf{z} = (z_1 \ldots z_n)$ are the observational variables measured from the origin.

The orthogonal mean regression is by definition a regression of the form (4.9) where the constant coefficients are determined by minimizing the function

$$\Sigma_t [g_0 + \mathbf{g} \cdot \mathbf{z}(t)]^2 / g^2$$

where $g = |\mathbf{g}| = +\sqrt{\mathbf{g}\mathbf{g}}$, and all the coefficients $g_0 g_1 \ldots g_n$ are considered as independent variables.

A derivation with respect to g_0 shows that we must have

$$\Sigma_t [g_0 + \mathbf{g} \cdot \mathbf{z}(t)] = 0.$$

Hence, the orthogonal mean regression goes through the mean of all observations. We can therefore at once simplify the problem by writing the regression

$$\mathbf{a}\mathbf{x} = 0 \tag{4.10}$$

where $\mathbf{x} = (x_1 \ldots x_n)$ are the variables measured from their respective means, and the coefficients $\mathbf{a} = (a_1 \ldots a_n)$ are to be determined by minimizing the function

$$\lambda = f(\mathbf{a}) = \Sigma_t [\Sigma_k a_k x_k(t)]^2 / a^2 = \frac{\mathbf{a}\mathbf{M}\mathbf{a}}{\mathbf{a}\mathbf{a}} \tag{4.11}$$

where $a = |\mathbf{a}| = +\sqrt{\mathbf{a}\mathbf{a}}$, and all the coefficients $a_1 \ldots a_n$ are considered as independent variables.

From (4.11) we get after a simple reduction

$$f_i = \frac{\partial f}{\partial a_i} = \frac{2}{a^2} \Sigma_k (m_{ik} - \lambda e_{ik}) a_k \tag{4.12}$$

where $e_{ij} = \begin{cases} 0 \ (i \neq j) \\ 1 \ (i = j) \end{cases}$ are the elements of the unit matrix $\mathbf{E}$, and λ has the value given by (4.11).

For the vector $\mathbf{f} = (f_1 \ldots f_n)$ we therefore have the expression

$$\frac{a^2}{2} \mathbf{f} = (\mathbf{M} - \lambda \mathbf{E})\mathbf{a}.$$

Now, the necessary condition for a minimum of $f(\mathbf{a})$ is $\mathbf{f} = 0$. The necessary conditions which $\mathbf{a}$ must satisfy, is therefore

$$(\mathbf{M} - \lambda \mathbf{E})\mathbf{a} = 0. \tag{4.13}$$

If there shall exist a solution of (4.13) other than the trivial $\mathbf{a} = 0$, the determinant $|\mathbf{M} - \lambda \mathbf{E}|$ must vanish. That is, λ must be one of the characteristic numbers of the moment matrix $\mathbf{M}$. This result regarding the regression coefficients $\mathbf{a}$ has first been obtained by Karl Pearson in his well know memoir "On lines and planes ..." (Phil. Mag, 1901).

If λ_i is any of the n characteristic numbers of $\mathbf{M}$, the corresponding vector $\mathbf{a}_i$ is simply determined by solving the linear system obtained from (4.13) by putting $\lambda = \lambda_i$. Inversely if $\mathbf{a}_i$ is any solution of (4.13) which is not $= 0$, the corresponding value of λ must necessarily be equal to $\lambda = \frac{\mathbf{a}_i \mathbf{M} \mathbf{a}_i}{\mathbf{a}_i \mathbf{a}_i}$. This simply follows from (4.13) by premultiplication with $\mathbf{a}_i$.

This being so, the essential question in the problem of the orthogonal mean regression is therefore: Will any of the characteristic numbers of $\mathbf{M}$, introduced in (4.13), furnish a solution $\mathbf{a}$ for which the function $f(\mathbf{a})$ actually attains a minimum? And, if so, which one (or which ones) of the characteristic numbers of $\mathbf{M}$ has this property?[1] The answer to this question is given by the following proposition.

If λ_1 is the smallest of the characteristic numbers of the moment matrix $\mathbf{M}$ (i. e. $\lambda_1 = 0$ if $\mathbf{M}$ is singular), and $\mathbf{a}_1$ is any solution of $(\mathbf{M} - \lambda_1 \mathbf{E})\mathbf{a} = 0$, which is not $= 0$, then the function $f(\mathbf{a}) = \frac{\mathbf{a}\mathbf{M}\mathbf{a}}{\mathbf{a}\mathbf{a}}$ is equal to λ_1 for $\mathbf{a} = \mathbf{a}_1$, and greater than λ_1 for any other argument $\mathbf{a}$ (not $= 0$) which is not a solution of $(\mathbf{M} - \lambda_1 \mathbf{E})\mathbf{a} = 0$.

[1] A rigorous distinction between maxima and minima is usually not made when the orthogonal mean regression is discussed. It is for instance generally assumed without further analysis that the extremum point where the value of $f(\mathbf{a})$ is less than in any other extremum point, is a minimum point of $f(\mathbf{a})$.

From (4.12) we get by derivation with respect to a_j

$$f_{ij} = \frac{\partial^2 f}{\partial a_i \partial a_j} = \frac{2}{a^2}[(m_{ij} - \lambda e_{ij}) - (a_i f_j + a_j f_i)].$$

For the matrix $\mathbf{F} = \begin{pmatrix} f_{11} \dots f_{1n} \\ \dots\dots\dots \\ f_{n1} \dots f_{nn} \end{pmatrix}$ of the second order partial derivatives, we therefore have in any point where $\mathbf{f} = 0$

$$\frac{a^2}{2}\mathbf{F} = (\mathbf{M} - \lambda\mathbf{E}). \tag{4.14}$$

Consequently: *A solution* $\mathbf{a}_i$ *(not* $= 0$*) of (4.13) for* $\lambda = \lambda_i$ *is a minimum point for* $f(\mathbf{a})$ *when and only when* $(\mathbf{M} - \lambda_i\mathbf{E})$ *is a positive definite matrix.* In particular we see that it will be a p-fold improper minimum point if $(\mathbf{M} - \lambda_i\mathbf{E})$ is positive definite and of rank $n - p$. The case where any minimum point is p-fold improper according to the criterion (4.14), is the very same case in which the totality of all minimum points form a p dimensional variety according to (4.13).

Since the rank of $(\mathbf{M} - \lambda_i\mathbf{E})$ is at most equal to $n - 1$, any minimum point of $f(\mathbf{a})$ must be at least one-fold improper. This is but another expression for the nature of our problem, as a problem in the direction, not in the length of $\mathbf{a}$. By the definition (4.11) a change in the length of $\mathbf{a}$ has no influence on the value of $f(\mathbf{a})$.

This being so, in order to prove our proposition it is sufficient to prove the lemma that $(\mathbf{M} - \lambda\mathbf{E})$ *is a positive definite matrix when and only when* $\lambda \leqq \lambda_1$, *where* λ_1 *is the smallest of the characteristic numbers of* $\mathbf{M}$.

From this will namely follow that among the points $\mathbf{a}$ (not $= 0$) where $\mathbf{f} = 0$, no other points than solutions of $(\mathbf{M} - \lambda_1\mathbf{E})\mathbf{a} = 0$ can give a minimum point for $f(\mathbf{a})$, and that these solutions actually do correspond to a minimum of $f(\mathbf{a})$.

Now, to prove the lemma let us perform on the set of observational variables $\mathbf{x}$ an orthogonal transformation $\mathbf{x}' = \mathbf{O}\mathbf{x}$ such that $\mathbf{x}'$ becomes an uncorrelated set, i. e. such that $\mathbf{M}' = \mathbf{O}\mathbf{M}\dot{\mathbf{O}}$ is a diagonal matrix $\mathbf{M}' = \begin{pmatrix} \lambda_1 \dots 0 \\ \dots\dots\dots \\ 0 \dots \lambda_n \end{pmatrix}$. From the theory of quadratic forms we know that such a transformation $\mathbf{O}$ always exists. Furthermore, the diagonal elements of $\mathbf{M}'$ will be the characteristic numbers of $\mathbf{M}$, hence all real. Since $\mathbf{O}$ is orthogonal, that is $\mathbf{O}\dot{\mathbf{O}} = \mathbf{E}$, we have

$$\mathbf{O}(\mathbf{M} - \lambda\mathbf{E})\dot{\mathbf{O}} = (\mathbf{O}\mathbf{M}\dot{\mathbf{O}} - \lambda\mathbf{O}\mathbf{E}\dot{\mathbf{O}}) = (\mathbf{M}' - \lambda\mathbf{E})$$

hence

$$\mathbf{O}(\mathbf{M} - \lambda\mathbf{E})\dot{\mathbf{O}} = \begin{pmatrix} \lambda_1 - \lambda, \dots 0 \\ \dots\dots\dots\dots \\ 0, \dots, \lambda_n - \lambda \end{pmatrix}. \tag{4.15}$$

A necessary and sufficient condition for the right hand side matrix in (4.15) to be positive definite, is that all the diagonal elements are non negative, i. e. $\lambda \leqq \lambda_1$, where λ_1 is the smallest of the characteristic numbers $\lambda_1 \dots \lambda_n$. But $\mathbf{O}(\mathbf{M} - \lambda\mathbf{E})\dot{\mathbf{O}}$ is positive definite when and only when $(\mathbf{M} - \lambda\mathbf{E})$ is positive definite. This simply follows from the fact that if $\mathbf{A}$ is any matrix, $\mathbf{C}\mathbf{A}\dot{\mathbf{C}}$ has the same rank and the same index as $\mathbf{A}$, provided only that $\mathbf{C}$ is non singular. Hence $(\mathbf{M} - \lambda\mathbf{E})$ is positive definite when and only when $\lambda \leqq \lambda_1$. This establishes our proposition. Incidentally, a similar argument shows that $(\mathbf{M} - \lambda\mathbf{E})$ is negative definite when and only when $\lambda_n \geqq \lambda$, where λ_n is the greatest of the characteristic numbers of $\mathbf{M}$. The argument evidently holds good without any assumption as to $\mathbf{M}$ being definite or not.

Summing up, we can state the following rule for determining the orthogonal mean regression: First find the smallest of the characteristic numbers for the moment matrix $\mathbf{M}$, that is the smallest of the n (necessarily real and non negative) roots of the secular equation $|\mathbf{M} - \lambda\mathbf{E}| = 0$. Let if be λ_1. Next solve the linear equation $(\mathbf{M} - \lambda_1\mathbf{E})\mathbf{a} = 0$ with respect to $\mathbf{a}$. The vector $\mathbf{a}$ being thus determined,

$$\mathbf{a}\mathbf{x} = 0 \tag{4.16}$$

is the orthogonal mean regression.

For the numerical computation of λ_1 the following limitation will be useful:

$$0 \leqq \lambda_1 \leqq m_{11} \tag{4.17}$$

where m_{11} is the smallest of the n sum squares $m_{ii} = \Sigma_t\,[x_i(t)]^2$. This formula is readily verified by noticing that m_{11} is one of the values which $f(\mathbf{a})$ can assume. In fact, for $a_2 = \dots = a_n = 0$

and a_1 arbitrary we have $f(\mathbf{a}) = m_{11}$. The minimum of $f(\mathbf{a})$, that is λ_1, can therefore never be greater than m_{11}. More generally: The same argument shows that if $\mathbf{M}_p$ is any p rowed principal submatrix in $\mathbf{M}$ $(1 \leqq p \leqq n)$, and λ_{1p} is the smallest characteristic number for $\mathbf{M}_p$, then we have the limitation

$$0 \leqq \lambda_1 \leqq \lambda_{1p}.$$

In the secular polynomial $|\mathbf{M} - \lambda \mathbf{E}|$ the coefficient of λ^{n-p} is equal to $(-)^{n-p} \Sigma M_p$, where ΣM_p is the sum of all the $\binom{n}{p}$ p-rowed principal minors in $\mathbf{M}$. Since all the characteristic numbers are non negative, we therefore also have the limitations

$$0 \leqq \binom{n}{p} \lambda_1^p \leqq \Sigma M_p \qquad (p = 1,2 \ldots n) \tag{4.18}$$

If the matrix $\mathbf{S} = (\mathbf{M} - \lambda_1 \mathbf{E})$ is exactly of rank $n-1$, the vector $\mathbf{a}$ is uniquely determined, apart from a scalar factor. In this case $\mathbf{a}$ is obtained by simply putting the components of $\mathbf{a}$ proportional to the elements of any row (not consisting exclusively of zeros) in the adjoint $\hat{\mathbf{S}}$ of $\mathbf{S}$. If the characteristic number λ_1 is computed exactly, all the rows of $\hat{\mathbf{S}}$ will be rigorously proportional, so it does not matter which row of $\hat{\mathbf{S}}$ is picked out. This might not hold good rigorously if λ_1 is determined by a numerical approximation method. In this case it will be better to determine only the signs of the components $a_1 \ldots a_n$ from the rows of $\hat{\mathbf{S}}$, and then determine the absolute values of $a_1 \ldots a_n$ by the formula

$$|a_i| = + \sqrt{\hat{s}_{ii}} \tag{4.19}$$

where $\hat{s}_{ii}$ is the i-th diagonal element in $\hat{\mathbf{S}}$, hence non negative, because $\mathbf{S}$ is positive definite. The formula (4.19) is proved thus. The rank of the symmetric and positive definite matrix $\mathbf{S}$ is never greater than $n-1$, and in the present case it is by assumption exactly $n-1$. Hence there is a diagonal element $\hat{s}_{kk}$ in $\hat{\mathbf{S}}$, which is positive, not zero. And the quantities a_i are proportional to the quantities $\hat{s}_{ki}$. But $\hat{s}_{ki}^2 = \hat{s}_{kk}\hat{s}_{ii}$ because $\hat{\mathbf{S}}$ is of rank 1. Since $\hat{s}_{kk} \neq 0$, we can therefore put $|a_i| = + \sqrt{s_{ii}}$.

The following is the explicit formula for the case $n = 2$. If

$m_{12} = m_{11} - m_{22} = 0$, that is if the two variables are uncorrelated and have the same standard deviation, the problem is not determinate. Any straight line through the mean of all observations will in this case satisfy the conditions for the orthogonal mean regression line. In any other case the orthogonal mean regression is uniquely determined and its equation is

$$\sqrt{p+q} \cdot x_1 - \varepsilon \sqrt{p-q} \cdot x_2 = 0 \tag{4.20}$$

where $p = \sqrt{m_{12}^2 + q^2}$ $\qquad q = (m_{22} - m_{11})/2$

and all the square roots are taken positive. Since the moment matrix is positive definite, we always have $p \geqq |q|$. ε is put equal to $+1$ or -1 according as m_{12} is positive or negative. If $m_{12} = 0$, it does not make any difference which one of the two values $+1$ and -1 is attributed to ε. From (4.20) follows in particular that if the variables are uncorrelated and have different standard deviations, the orthogonal mean regression is a straight line parallel to the axis of the variable which has the largest standard deviation.

It is readily seen that the mean regression determined by minimizing the sum square of the deviations measured under a certain fixed angle $(\neq 0)$ with the regression plane, coincides with the orthogonal mean regression, regardless of the magnitude of the fixed angle. But a regression determined by minimizing the sum square of the deviations measured under certain fixed angles with the axes of the variables, will depend on the magnitude of these angles.

I shall now consider a mean regression which is not determined directly by least squares, but by a principle more similar to the principle of formula (4.19).

The expression (4.6) for the regression matrix was developped under the assumption that $\mathbf{M}$ is non singular. The expression (4.6) still holds good, however, in the case where $\mathbf{M}$ is of rank $n-1$. In fact, in this case there exists identically in t at least one linear relation of the form (1.3). Multiplying (1.3) by x_j and performing a summation over t, we see that the vector $\mathbf{a} = (a_1 \ldots a_n)$ must be a solution of

$$\mathbf{M}\mathbf{a} = 0. \tag{4.21}$$

If $\mathbf{M}$ is of rank $n-1$, the equation (4.21) has a solution $\mathbf{a}$

which is determined uniquely, except for an arbitrary scalar factor. Apart from an arbitrary factor of proportionality, only one equation of the form (1.3) can therefore exist when $\mathbf{M}$ is of rank $n-1$. This equation which holds good identically in t when $\mathbf{M}$ rank $n-1$, might be called *the perfect regression.*

The solution $\mathbf{a}$ of (4. 21) is obtained by putting the components $a_1 \ldots a_n$ proportional to the elements in any row of the adjoint moment matrix $\hat{\mathbf{M}}$, which is such that the diagonal element in this row is different from zero. At least one such row exists since the symmetric matrix $\hat{\mathbf{M}}$ is of rank 1 when $\mathbf{M}$ is of rank $n-1$.

But this is exactly the system of regression coefficients which are furnished by the formula (4.6). For if $\hat{\mathbf{M}}$ is of rank 1, all the rows of $\hat{\mathbf{M}}$ are proportional to one of them, say the k-th, which is such that $\hat{m}_{kk} \neq 0$. All the regression equations determined by (4.6) will therefore coincide and coincide with the perfect regression. Some of the regressions furnished by (4.6) may however be trivial, all the regression coefficients being equal to zero. That is, some of the rows of $\hat{\mathbf{M}}$ (not the k-th, however) might consist exclusively of zeros. These rows can be looked upon as being obtained by multiplying the k-th row by 0.

When $\hat{\mathbf{M}}$ is of rank 1 we have $\hat{m}_{kj}^2 = \hat{m}_{kk}\hat{m}_{jj}$ $(j = 1,2 \ldots n)$, where $\hat{m}_{kk} \neq 0$. The perfect regression which exists when $\mathbf{M}$ is of rank $n-1$, can therefore be written

$$\Sigma_i \varepsilon_i \sqrt{\hat{m}_{ii}} \cdot x_i = 0 \tag{4.22}$$

or in standard coordinates $y_i = x_i/\sigma_i$.

$$\Sigma_i \varepsilon_i \sqrt{\hat{r}_{ii}} \cdot y_i = 0 \tag{4.23}$$

where the square roots are taken positive, and $\varepsilon_i = \pm 1$ is the sign of the i-th element in the k-th row of $\hat{\mathbf{M}}$, k can be chosen arbitrarily provided $\hat{m}_{kk} \neq 0$. The formula (4.23) shows the significance of $\sqrt{\hat{r}_{ii}}$ as a coefficient of linear importance for the variable x_i.

Now, if the signs in the rows of $\hat{\mathbf{M}}$ are compatible, the equation (4.22), respectively (4. 23), is a well defined equation, which can be computed even though $\mathbf{M}$ is non singular. In the case of compatible signs in the rows of $\hat{\mathbf{M}}$, the equation (4.22), respectively (4.23), may therefore be used as a mean regression. Since

irs coefficients are derived from the diagonal elements of $\hat{\mathbf{M}}$, tespectively of $\hat{\mathbf{R}}$, it might be called *the diagonal mean regression.*

In connection with the regression equations I shall make a few remarks on the partial and multiple correlation coefficients. The formula (2.1) shows that the normalized of the moment matrix $\mathbf{M}$ is the matrix $\mathbf{R}$ of the simple correlation coefficients r_{ij}. Similarly the normalized of the adjoint moment matrix $\hat{\mathbf{M}}$ (with the sign reversed) is the matrix $\bar{\mathbf{R}}$ of the highest order partial correlation coefficients $\bar{r}_{ij}$ (in professor *Yule's* notation $r_{ij.12\ldots n}$ where ij is not written in the string of secondary subscripts).

The partial correlation coefficients $\bar{r}_{ij}$ are sometimes expressed by the formula

$$\bar{r}_{ij} = \sqrt{b_{ij}\, b_{ji}}$$

where b_{ij} are the elements of the reduced regression matrix $\mathbf{B}$. In this form the expression is however incomplete. In order to make it complete we have to show that b_{ij} and b_{ji}, are always of the same sign, $sgn\, b_{ij} = sgn\, b_{ji}$, and then write the formula

$$\bar{r}_{ij} = sgn\, b_{ij} \sqrt{b_{ij}\, b_{ji}} \tag{4.24}$$

where the square root is taken positive.

That $sgn\, b_{ij} = sgn\, b_{ji}$ follows from (4.5) because $\mathbf{A}$ is symmetric and all the a_{ii} are positive, not zero, when $\mathbf{M}$ is non singular (which has to be assumed when the partial correlation coefficients $\bar{r}_{ij}$ are considered).

The formula (4.24) is somewhat artificial. The formula becomes simpler and more rational when expressed in terms of the regression coefficients a_{ij}. Since $sgn\, b_{ij} = -sgn\, a_{ij}$, we have

$$\bar{r}_{ij} = -\frac{a_{ij}}{\sqrt{a_{ii}\, a_{jj}}} = -\frac{\hat{m}_{ij}}{\sqrt{\hat{m}_{ii}\, \hat{m}_{jj}}} \tag{4.25}$$

where the square roots are taken positive.

If we prefer, we can also express $-\bar{\mathbf{R}}$ as the normalized of $\hat{\mathbf{R}}$. In fact the normalized of $\hat{\mathbf{M}}$ must be equal to the normalized of $\hat{\mathbf{R}}$, for by (2.10) there exists a diagonal matrix $\mathbf{D}'$ with positive diagonal elements, such that $\hat{\mathbf{R}} = \mathbf{D}'\hat{\mathbf{M}}\mathbf{D}'$. Hence

$$\bar{r}_{ij} = -\frac{\hat{r}_{ij}}{\sqrt{\hat{r}_{ii}\, \hat{r}_{jj}}}. \tag{4.26}$$

In numerical work a uniform exactitude to a given number of decimals is more easily obtained from (4.26) than from (4.25).

In exactly the same way the lower order partial correlation coefficients can be expressed by normalizing the adjoints of the various principal submatrices in $\hat{\mathbf{R}}$.

For the multiple correlation coefficients r_i (in Professor *Yule's* notation $R_{i(12\ldots n)}$ where i is not written in the string of secondary subscripts) we have

$$r_i = +\sqrt{1-\frac{R}{\hat{r}_{ii}}}. \tag{4.27}$$

There is a certain analogy between this formula and the formula

$$r = +\sqrt{1-R}$$

for the collective correlation coefficient. The collective, multiple and partial correlation coefficients can be looked upon as forming a sort of hierarchic order. The number of their subscripts is 0, 1, 2.

The absolute value of an element in $\overline{\mathbf{R}}$ is never less than the absolute value of the corresponding element in $\hat{\mathbf{R}}$, for the elements in $\overline{\mathbf{R}}$ are obtained by dividing the elements of $\hat{\mathbf{R}}$ by the quantities $\sqrt{\hat{r}_{ii}\hat{r}_{jj}}$, and these quantities are never greater than unity. This follows from the fact that each $\hat{r}_{ii}$ is itself the square of a scatter coefficient (namely the scatter coefficient for the set obtained from $\mathbf{x}$ by omitting x_i).

It is further easily seen that not only is the absolute value of each element in $\hat{\mathbf{R}}$ less than or equal to unity but even the absolute value of each element in $\overline{\mathbf{R}}$ is less than or equal to unity. For $\mathbf{R}$ (and hence $\hat{\mathbf{R}}$) is positive definite. Therefore all the principial minors of $\hat{\mathbf{R}}$ and particularly all the two rowed principal minors of $\hat{\mathbf{R}}$ are non negative. That is

$$\hat{r}_{ii}\hat{r}_{jj} \geqq \hat{r}_{ij}^2 \qquad (i,j=1,2\ldots n)$$

We therefore have

$$0 \leqq |\hat{r}_{ij}| \leqq |\overline{r}_{ij}| \leqq 1 \qquad (i,j=1,2\ldots n). \tag{4.28}$$

The preceding formulae for the multiple and partial correlation coefficients only hold good if $R \neq 0$. In this case all the $\hat{r}_{ii}$ are positive, not zero, because they are principal minors in a positive definite and non singular matrix. The case $R=0$ is discussed in Sections 6 and 7.

5. THE PROBLEM OF INVARIANCE.

I shall now consider the effect on the regression equations which is produced by a linear transformation performed on the observational variables.

Let $\mathbf{z}=(z_1\ldots z_n)$ be the observational variables measured from the origin. If $\mathbf{c}$ is an arbitrary vector and $\mathbf{C}$ a non singular matrix, the general non singular linear transformation may be written

$$\mathbf{z}' = \mathbf{c} + \mathbf{C}\mathbf{z}. \tag{5.1}$$

The following are some special cases:

A *translation*, i. e. a change of origin, is the transformation $\mathbf{z}'=\mathbf{z}+\mathbf{c}$, obtained from (5.1) by puttig $\mathbf{C}=\mathbf{E}$.

A *stretch*, i. e. a change of scales, is the transformation $\mathbf{z}'=\mathbf{D}\mathbf{z}$, where $\mathbf{D}$ is a diagonal matrix.

A *distance preserving transformation* is the transformation $\mathbf{z}'=\mathbf{O}\mathbf{z}$ where $\mathbf{O}$ is an orthogonal matrix.

A *homogeneous transformation* is the transformation $\mathbf{z}'=\mathbf{C}\mathbf{z}$, obtained by putting $\mathbf{c}=0$ in (5.1). $\mathbf{C}$ is assumed non singular but otherwise arbitrary. A homogeneous linear transformation includes all types of linear transformations except translations.

The equation of any regression can be written in the form $g_0+\mathbf{g}\mathbf{z}=0$, where g_0 and $\mathbf{g}$ are determined from the totality of the observations by a certain rule or law, which is characteristic for the regression method in question. If we perform on the observational set $\mathbf{z}$ a transformation of the form (5.1), the set $\mathbf{z}'$ will also be an observational set, in which we may determine the regression by the rules of the regression method in question. Let the regression in the set $\mathbf{z}'$ be $g_0'+\mathbf{g}'\mathbf{z}'=0$. If we introduce in this equation the expression (5.1) for $\mathbf{z}'$ in terms of $\mathbf{z}$, we obtain a regression $g_0''+\mathbf{g}''\mathbf{z}=0$ which now involves the original set $\mathbf{z}$. This equation might be called the regression in $\mathbf{z}$ determined via $\mathbf{z}'$. If the equation $g_0''+\mathbf{g}''\mathbf{z}=0$ is the same as the equation $g_0+\mathbf{g}\mathbf{z}=0$, that is, if the coefficients of the two equations are proportional, then the particular type of regression considered is called *invariant* under the transformation (5.1). In this

case it does not make any difference if the regression in $\mathbf{z}$ is determined directly or via $\mathbf{z}'$.

It is readily seen that the elementary regressions (4.7) are invariant for a translation. Both the moment matrix $\mathbf{M}$ and the set $\mathbf{x}$ of variables measured from their respective means, are namely unchanged under a translation.

If a homogeneous linear transformation $\mathbf{z}' = \mathbf{Cz}$ is performed, the regression matrix $\mathbf{A}$ is replaced by

$$\mathbf{A}' = \check{\mathbf{C}}\mathbf{A}\hat{\mathbf{C}} = C^2\grave{\mathbf{C}}^{-1}\mathbf{A}\mathbf{C}^{-1}. \tag{5.2}$$

In fact, from $\mathbf{z}' = \mathbf{Cz}$ we get $\mathbf{x}' = \mathbf{Cx}$ and therefore $\mathbf{M}' = \mathbf{CM}\grave{\mathbf{C}}$. Taking the adjoint of this equation we get (5.2). Formula (5.2) can be stated thus: Apart from a constant factor (which is the square of the modulus of transformation) the regression matrix $\mathbf{A}$ is transformed as the matrix in the quadratic form $\mathbf{xAx}$, where the variables are cogredient with the observational variables.

Now, let us go back from $\mathbf{x}'$ to the original set $\mathbf{x}$, this time performing the transformation on the variables in the regression equations. Introducing $\mathbf{x}' = \mathbf{Cx}$ in the regression system $\check{\mathbf{C}}\mathbf{A}\hat{\mathbf{C}}\mathbf{x}' = 0$, we get

$$\check{\mathbf{C}}\mathbf{Ax} = 0 \tag{5.3}$$

which is the system of regression equations in $\mathbf{x}$ determined via $\mathbf{x}' = \mathbf{Cx}$. This system is equivalent with the system $\mathbf{Ax} = 0$ if $\mathbf{C}$ is a diagonal matrix, in which case the transformation $\mathbf{x}' = \mathbf{Cx}$ is a stretch, which simply means changing the scales of the variables. If $\mathbf{C}$ is not diagonal, the new regression system will, in general, be different from the original (and it will certainly be different if $\mathbf{M}$ is non singular). The formula (5.3) even shows that by a suitable choice of the set $\mathbf{x}'$ via which the regression system in $\mathbf{x}$ is determined, we can get any regression system in $\mathbf{x}$ we want. In fact, if $\mathbf{Ax} = 0$ is the regression system in $\mathbf{x}$ determined directly, and we want to obtain the regression system $\mathbf{Px} = 0$, we only have to determine the regression system in $\mathbf{x}$ via $\mathbf{x}' = \check{\mathbf{P}}\grave{\mathbf{A}}\mathbf{x}$.

The orthogonal mean regression is invariant for a translation and also *for a general orthogonal transformation.* That it is invariant for a translation simply follows from the fact that it goes through the mean of all observations and the coefficient vector $\mathbf{a}$ in (4.16) is determined by the elements of the moment matrix $\mathbf{M}$, and these are not changed by a translation.

That the orthogonal mean regression is invariant for an orthogonal transformation, is proved thus. Let $\mathbf{x}' = \mathbf{Ox}$, hence $\mathbf{M}' = \mathbf{OM}\grave{\mathbf{O}}$ and consequently $\mathbf{O}(\mathbf{M} - \lambda\mathbf{E})\grave{\mathbf{O}} = (\mathbf{M}' - \lambda\mathbf{E})$, since $\mathbf{O}\grave{\mathbf{O}} = \mathbf{E}$. By taking the determinant on both sides, we see that $|\mathbf{M} - \lambda\mathbf{E}| = |\mathbf{M}' - \lambda\mathbf{E}|$ for any value of λ. The secular polynomial for $\mathbf{M}'$ is therefore identical with the secular polynomial for $\mathbf{M}$. The set of n characteristic numbers will therefore coincide for $\mathbf{M}'$ and $\mathbf{M}$, in particular the smallest of the characteristic numbers, namely λ_1, will be the same for $\mathbf{M}'$ and $\mathbf{M}$.

Now, the coefficient vectors $\mathbf{a}$ and $\mathbf{a}'$ for the orthogonal mean regressions in the set $\mathbf{x}$ and $\mathbf{x}'$ respectively are determined by the equations

$$(\mathbf{M} - \lambda_1\mathbf{E})\mathbf{a} = 0 \text{ and } (\mathbf{M}' - \lambda_1\mathbf{E})\mathbf{a}' = 0.$$

The equation for $\mathbf{a}'$ can be written $\mathbf{O}(\mathbf{M} - \lambda_1\mathbf{E})\grave{\mathbf{O}}\mathbf{a}' = 0$. Since $\mathbf{O}$ is non singular, this is equivalent with $(\mathbf{M} - \lambda_1\mathbf{E})\grave{\mathbf{O}}\mathbf{a}' = 0$. $\grave{\mathbf{O}}\mathbf{a}'$ and $\mathbf{a}$ are thus solutions of the same equation. If $\mathbf{a}$ runs through all solutions of the equation for $\mathbf{a}$, the vector $\mathbf{Oa}$ will run through all solutions of the equation for $\mathbf{a}'$. If we let $\mathbf{a}$ and $\mathbf{a}'$ denote general solutions of the two equations above, we therefore have $\mathbf{a}' = \mathbf{Oa}$. Now, $\mathbf{a}'\mathbf{x}' = 0$ is the orthogonal mean regression in the set $\mathbf{x}'$. Introducing in this equation the expression for $\mathbf{x}'$, namely $\mathbf{x}' = \mathbf{Ox}$, we get $(\mathbf{Oa})(\mathbf{Ox}) = 0$, hence $\mathbf{a}\grave{\mathbf{O}}\mathbf{Ox} = 0$, that is $\mathbf{ax} = 0$, which is the orthogonal mean regression determined directly in the set $\mathbf{x}$.

The diagonal mean regression (4.22) has the property of being invariant for translation and stretch but not for a general linear transformation.

The fact that the regressions here considered are only invariant under rather special forms of transformations, gives rise to a serious difficulty.

In a great number of cases, we have no absolute *a priori* criterion to guide us in the choice of variables. This is particularly true in the case of a differential analysis of time series. For instance: Shall we determine a regression between successive differences or shall we determine a regression between consecutive ordinates? In fact, it was precisely this problem which led me to consider linear correlation from the point of view of linear transformations.

So far as practical application is concerned, it is not in all cases necessary (and it might not even be desirable) to introduce

a mean regression plane which is invariant for an absolutely general linear transformation. In some cases the *origin* of the set of variables contained in the original data might be in a sense fixed by the nature of the problem, so that it is not particularly necessary that the mean regression plane should be invariant for a translation (i. e. for a change of origin). What we want in this type of problem is a mean regression plane which is invariant for a *homogeneous* linear transformation of the variables measured from the origin. I now proceed to an analysis of this problem.

In the case of a homogeneous linear transformation, the variables measured from the origin, i. e. the set $\mathbf{z}$, and the variables measured from the means, i. e. $\mathbf{x}$, are cogredient. In fact, if $\mathbf{z}' = \mathbf{Cz}$, then $\mathbf{x}' = \mathbf{Cx}$. The inverse also holds good if the convention is made that under a homogeneous linear transformation of $\mathbf{x}$, the vector whose components are the means of the variables, shall be cogredient with $\mathbf{x}$. This convention being adopted, the expressions "a homogeneous linear transformation of the observational variables" and "a vector cogredient with the observational variables" are unambigeous without specifying if $\mathbf{x}$ or $\mathbf{z}$ is meant.

Let $k_i = \Sigma_t z_i(t)$ and $\mathbf{k} = (k_1 \dots k_n)$, so that $\mathbf{k}/\omega$ is the vector whose components are the means of the variables. Further let $k_{ij} = \Sigma_t z_i(t) z_j(t)$, so that

$$\mathbf{K} = \begin{pmatrix} k_{11} \dots k_{1n} \\ \dots\dots\dots \\ k_{n1} \dots k_{nn} \end{pmatrix}$$

is the moment matrix taken about the origin. $K = |\mathbf{K}|$ is the determinant value of $\mathbf{K}$.

If μ and χ are arbitrary constants (not necessarily $\neq 0$) and $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ and $\mathbf{v}$ are vectors cogredient with the observational variables, but otherwise arbitrary, then the two scalar quantities

$$\mu M + \chi K$$

and

$$\mathbf{p\hat{M}q} + \mathbf{u\hat{K}v}$$

are both covariants of weight two. That is, if a homogeneous linear transformation is performed on the observational variables, the only effect on the two scalars considered, is that they are multiplied by the square of the modulus of transformation. In fact, if we perform the transformation $\mathbf{z}' = \mathbf{Cz}$, that is $\mathbf{x}' = \mathbf{Cx}$,

we have $\mathbf{M}' = \mathbf{CM\dot{C}}$ and $\mathbf{K}' = \mathbf{CK\dot{C}}$, consequently $\mathbf{\hat{M}}' = \mathbf{\check{C}\hat{M}\hat{C}}$ and $\mathbf{\hat{K}}' = \mathbf{\check{C}\hat{K}\hat{C}}$. Hence

$$\mu M' + \chi K' = C^2(\mu M + \chi K) \tag{5.4}$$

and

$$\mathbf{p'\hat{M}'q'} + \mathbf{u'\hat{K}'v'} = \mathbf{p\check{C}\check{C}\hat{M}\hat{C}\hat{C}q} + \mathbf{u\check{C}\check{C}\hat{K}\hat{C}\hat{C}v} = C^2(\mathbf{p\hat{M}q} + \mathbf{u\hat{K}v}). \tag{5.5}$$

The equation of a regression plane for the variables $\mathbf{z}$ can be written

$$g_0 + \mathbf{gz} = 0 \tag{5.6}$$

where $\mathbf{g} = (g_1 \dots g_n)$. The coefficients $g_0 g_1 \dots g_n$ are constants independent of t, but depending on the totality of the observations. In order to indicate this explicitly we use the notation

$$g_0 = g_0(\mathbf{z}) \qquad \mathbf{g} = \mathbf{g}(\mathbf{z}).$$

The regression $g_0 + \mathbf{gz} = 0$ *is invariant for a homogeneous linear transformation of the observational variables when, and only when* g_0 *and* $\mathbf{g}$ *are such that for an arbitrary non singular matrix* $\mathbf{C}$.

$$g_0(\mathbf{Cz}) = \gamma\, g_0(\mathbf{z}) \text{ and } \mathbf{g}(\mathbf{Cz}) = \gamma \mathbf{\dot{C}}^{-1} \cdot \mathbf{g}(\mathbf{z}) \tag{5.7}$$

where $\gamma \neq 0$ *is a scalar* (which may depend on the observations and on the elements of $\mathbf{C}$).

In fact, if the transformation $\mathbf{z}' = \mathbf{Cz}$ is performed on the set of observational variables $\mathbf{z}$ with the moment matrix $\mathbf{K}$, the moment matrix for the set $\mathbf{z}'$ will be $\mathbf{K}' = \mathbf{CK\dot{C}}$. If g_0 and $\mathbf{g}$ are of the form stated, the mean regression (determined directly) in the set $\mathbf{z}'$, namely $g_0(\mathbf{z}') + \mathbf{g}(\mathbf{z}') \cdot \mathbf{z}' = 0$, can be written $g_0(\mathbf{z}) + \mathbf{g}(\mathbf{z}) \cdot \mathbf{C}^{-1}\mathbf{z}' = 0$. Going back to the variables $\mathbf{z}$, now performing the transformation on the variables in the regression for $\mathbf{z}'$ we consequently obtain $g_0(\mathbf{z}) + \mathbf{g}(\mathbf{z}) \cdot \mathbf{z} = 0$, which is the mean regression determined directly in the set $\mathbf{z}$. Inversely it is readily seen that the specified conditions are also necessary for the invariance of the mean regression plane.

If χ is an arbitrary constant (not necessarily $\neq 0$), and $\mathbf{p}$,$\mathbf{q}$ and $\mathbf{r}$ are vectors which are cogradient with the observational variables, but otherwise arbitrary, then

(5.8) $$g_0 = \chi K \qquad \mathbf{g} = \mathbf{r}\hat{\mathbf{K}}$$

and

(5.9) $$g_0 = \mathbf{p}\hat{\mathbf{K}}\mathbf{q} \qquad \mathbf{g} = \mathbf{r}\hat{\mathbf{K}}$$

are two forms of the set $(g_0\mathbf{g})$ which satisfy the invariance condition (5.7). If $\mathbf{z}' = \mathbf{C}\mathbf{z}$, we namely have $\mathbf{g}' = \mathbf{r}'\hat{\mathbf{K}}' = \mathbf{r}\hat{\mathbf{C}}\hat{\mathbf{C}}\hat{\mathbf{K}}\hat{\mathbf{C}} = C^2\,\hat{\mathbf{C}}^{-1}\cdot(\mathbf{r}\hat{\mathbf{K}}) = C^2\hat{\mathbf{C}}^{-1}\mathbf{g}$, and furthermore by (5.4) and (5.5) $g_0' = C^2 g_0$.

If the relation (5.6) shall be, not only invariant, but also such that it can be considered the equation of a regression plane, the coefficients must be determined by some kind of fitting procedure.

I shall now show how a regression plane with coefficients of the form (5.8) and another mean regression plane with coefficients of the form (5.9) can be constructed by least square fitting procedures in the case where not all the variables (measured from the origin) have a mean equal to zero.

We first notice the following useful relation between the determinants M and K and the algebraic value of the quadratic forms $\mathbf{k}\hat{\mathbf{M}}\mathbf{k}$ and $\mathbf{k}\hat{\mathbf{K}}\mathbf{k}$.

(5.10) $$\omega M = \omega \begin{vmatrix} m_{11} \dots m_{1n} \\ \dots\dots\dots \\ m_{1n} \dots m_{nn} \end{vmatrix} = \begin{vmatrix} \omega & k_1 & \dots & k_n \\ k_1 & k_{11} & \dots & k_{1n} \\ \dots & \dots & \dots & \dots \\ k_n & k_{n1} & \dots & k_{nn} \end{vmatrix} = \omega K - \mathbf{k}\hat{\mathbf{K}}\mathbf{k}$$

(5.11) $$\omega K = \omega \begin{vmatrix} k_{11} \dots k_{1n} \\ \dots\dots\dots \\ k_{n1} \dots k_{nn} \end{vmatrix} = -\begin{vmatrix} -\omega & k_1 & \dots & k_n \\ k_1 & m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n & m_{n1} & \dots & m_{nn} \end{vmatrix} = \omega M + \mathbf{k}\hat{\mathbf{M}}\mathbf{k}$$

and similary for any subset of q variables, i. e. for any set of q subscripts picked from the set $1,2 \dots n$ (5.10) is proved thus: In the right hand side determinant subtract from the second row the first row multiplied by k_1/ω, next subtract from the third row the first row multiplied by k_2/ω, and so on. By virtue of the formula

$$k_{ij} = m_{ij} + k_i k_j/\omega$$

the determinant thus obtained will be

$$\begin{vmatrix} \omega & 0 & \dots & 0 \\ k_1 & m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n & m_{n1} & \dots & m_{nn} \end{vmatrix} = \omega M$$

(5.11) is proved in a similar way. The formulæ (5.10) and (5.11) can be looked upon as the generalization to n variables of the formulæ

$$\omega M = \omega K - k \cdot 1 \cdot k$$
$$\omega K = \omega M + k \cdot 1 \cdot k$$

which hold good for a single variable whose mean is k/ω and whose sum square about the mean and about the origin is M and K respectively. (5.10) and (5.11) can also be written

(5.12) $$\omega(K - M) = \mathbf{k}\hat{\mathbf{K}}\mathbf{k} = \mathbf{k}\hat{\mathbf{M}}\mathbf{k}.$$

The value of K (and its principal minors) is given by the formula obtained from (1.2) by changing M to K and x to z. Therefore, K is always non negative. And K is equal to zero when and only when there exists identically in t a relation of the form.

(5.13) $$g_1 z_1(t) + \dots + g_n z_n(t) = 0$$

where the coefficients g_i are independent of t and not all equal to zero. If such a relation holds good, we get by performing a summation over t

$$g_1 k_1 + \dots + g_n k_n = 0$$

and consequently

$$g_1 x_1(t) + \dots + g_n x_n(t) = 0.$$

If K is equal to zero, M must therefore also be equal to zero (although the inverse is not true). This fact also follows from (5.12) by noticing that $\mathbf{k}\hat{\mathbf{K}}\mathbf{k}$ is a positive definite quadratic form, and therefore

$$0 \leqq M \leqq K.$$

By an argument similar to that used in establishing (5.10) we obtain the formulæ

(5.14) $$M - \mathbf{k}\hat{\mathbf{M}}\mathbf{x} = \begin{vmatrix} 1 & x_1 & \dots & x_n \\ k_1 & m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n & m_{n1} & \dots & m_{nn} \end{vmatrix} = \begin{vmatrix} 1 & z_1 & \dots & z_n \\ k_1 & k_{11} & \dots & k_{1n} \\ \dots & \dots & \dots & \dots \\ k_n & k_{n1} & \dots & k_{nn} \end{vmatrix} = K - \mathbf{k}\hat{\mathbf{K}}\mathbf{z}$$

and

$$\begin{vmatrix} 0 & v_1 & \dots & v_n \\ k_1 m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n m_{n1} & \dots & m_{nn} \end{vmatrix} = \begin{vmatrix} 0 & v_1 & \dots & v_n \\ k_1 k_{11} & \dots & k_{n1} \\ \dots & \dots & \dots & \dots \\ k_n k_{n1} & \dots & k_{nn} \end{vmatrix}$$

that is

$$\mathbf{k}\hat{\mathbf{K}}\mathbf{v} = \mathbf{k}\hat{\mathbf{M}}\mathbf{v} \tag{5.15}$$

where $\mathbf{v} = (v_1 \dots v_n)$ is an arbitrary vector. Since $\mathbf{v}$ is arbitrary, we must further have

$$\mathbf{k}(\hat{\mathbf{K}} - \hat{\mathbf{M}}) = 0. \tag{5.16}$$

Incidentally, (5.16) shows that we always have

$$|\hat{\mathbf{K}} - \hat{\mathbf{M}}| = 0$$

and similarly for any subset of q subscripts ($1 \leqq q \leqq n$) picked from the set $1,2 \dots n$.

Now, a regression with coefficients of the form (5.8) is obtained simply by minimizing the sum square

$$\Sigma_t [g_0 + \mathbf{g} \cdot \mathbf{z}(t)]^2$$

where $g_0 \neq 0$ is considered as an arbitrary fixed constant (not to be determined by the minimizing conditions). This leads to the equation in $\mathbf{g}$

$$g_0 \mathbf{k} + \mathbf{g}\mathbf{K} = 0. \tag{5.17}$$

We may assume $K \neq 0$, otherwise there would exist identically in t a relation of the form (5.13) and the problem of a regression plane would be trivial. We can therefore solve for $\mathbf{g}$, which gives the regression

$$K - \mathbf{k}\hat{\mathbf{K}}\mathbf{z} = 0. \tag{5.18}$$

This is a regression with coefficients satisfying (5.8). In fact, if no translation is involved, $\mathbf{k}$ will be cogredient with the observational variables. The regression (5.18) is therefore invariant for a homogeneous linear transformation of the observational variables.

By (5.14) the equation (5.18) can also be written

$$\mathbf{k}\hat{\mathbf{M}}\mathbf{x} = M \tag{5.19}$$

or in determinant form

$$\begin{vmatrix} 1 & x_1 & \dots & x_n \\ k_1 m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n m_{n1} & \dots & m_{nn} \end{vmatrix} = 0. \tag{5.20}$$

We may improve on (5.18) by writing it

$$H - \mathbf{k}\hat{\mathbf{K}}\mathbf{z} = 0$$

and determining H by the least square condition

$$\Sigma_t [H - \mathbf{k}\hat{\mathbf{K}} \cdot \mathbf{z}(t)]^2 \text{ a minimum.}$$

This gives the regression

$$\mathbf{k}\hat{\mathbf{K}}\mathbf{k}/\omega - \mathbf{k}\hat{\mathbf{K}}\mathbf{z} = 0 \tag{5.21}$$

the coefficients of which are of the form (5.9). The regression (5.21) is therefore also invariant for a homogeneous linear transformation of the observational variables.

By virtue of (5.15) and (5.16) the equation (5.21) reduces to

$$\mathbf{k}\hat{\mathbf{M}}\mathbf{x} = 0 \tag{5.22}$$

or in determinant form

$$\begin{vmatrix} 0 & x_1 & \dots & x_n \\ k_1 m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ k_n m_{n1} & \dots & m_{nn} \end{vmatrix} = 0. \tag{5.23}$$

It is interesting to notice that the invariant regression (5.23) can be looked upon as the regression obtained by taking the weighted arithmetic average of the coefficients in the n elementary regressions, the weights being the n means of the variables, (5.23) might therefore be called the *composite mean regression.* Generally (5.23) will give a better fit to the observations than (5.20). The regression (5.23) goes through the mean of all observations, while this is not true of (5.20)[1].

[1] After having constructed the type of invariant regression which is considered in the present section, I learned from Dr. *C. F. Roos* of the important, not yet published, work which he and Dr. *A. Oppenheim* have been doing on invariant regressions. Roos and Oppenheim consider the case where the

6. COLLINEAR SETS AND PERFECT CORRELATION.

I now proceed to a discussion of the extreme cases of linear dependency and perfect correlation between statistical variables. It is true that these extreme cases are not the ones which are most likely to occur in practical applications. Nevertheless, as I see it, a discussion of these extreme cases is essential for a rational interpretation of what is really involved in the notion of correlation in several variables, and particularly for the interpretation of partial and multiple correlation coefficients which are close to unity.

For the simple correlation coefficient the case is clear enough. The simple correlation coefficient between two variables is equal to unity when and only when one variable is proportional to the other (each variable being measured from its mean). Geometrically interpreted: The simple correlation coefficient is close to unity when and only when the swarm of observation points in the two dimensional scatter diagram is clustering around a straight line through the origin. For partial and multiple correlation coefficients, however, the case is not quite so clear.

For more than two variables we have to distinguish between different *types of clustering*. For three variables for instance it might happen that the swarm of observation points in the three dimensional scatter diagram is clustering around a plane through the origin, but are highly scattered within this plane[1]. The plane

regression coefficients are determined by minimizing a function of the type $\Sigma_i f(\mathrm{z})$ where $f(\mathrm{z}) = f(z_1 \ldots z_n)$ is a function of the observational variables. The minimizing process is considered in the form where the regression cofficients by definition are imposed to satisfy a side relation. One of their beautiful results is that if the regression shall be invariant, the necessary and sufficient condition which f must satisfy, is $f(\mathrm{z}) = |g_0 + \mathrm{gz}|^\alpha$ where α is an arbitrary exponent and $g_0 g_1 \ldots g_n$ are the regression coefficients. Furthermore, the side relation must of course also be invariant. The Roos-Oppenheim criterion will certainly prove to be a most powerful tool in the research of new types of invariant regressions.

At the time of my conversation with Dr. Roos, the authors had not yet actually constructed any invariant regression. The main difficulty in doing so will probably be to construct invariant side relations. A type of side relation wich shall lead to regressions other than those contained in the formulae (5.8) and (5.9) above, will probably not be very simple.

[1] For the sake of brevity I use the expression »within the plane» allthough not all the points (perhaps none of them) are rigorously lying in the plane. The meaning is that the points are highly scattered in directions parallel to the plane.

may be far from containing any of the axes. Or it may be a plane containing say the x_1 axis, that is a plane perpendicular to the $x_2 x_3$ plane. Again it might happen that the swarm of observation points is clustering not only around a plane but even around a straight line in this plane. For several variables the number of different cases is of course much greater. *And each one of these various cases has a very definite significance.*

A rigorous analysis of these various cases is however usually neglected. And the practical application of the theory is often limited to the mechanical use of computation formulae under some simple assumptions, for instance the assumption that the partial correlation coefficient for the pair of variables $x_i x_j$ means the same thing as the simple correlation coefficient for $x_i x_j$ would have meant if the material had been, not what it actually is, but a material where all the other variables had been constants. (Which assumption is reasonably plausible only if the distribution of the variables is normal.)

The reason for not analyzing the various types of clustering by the classical methods of correlation is easily understood. It will presently appear that the partial and multiple correlation coefficients and other classical correlation parameters become undefined in those cases of linear dependency which illustrate the various types of clustering. And in practical cases which approache these extreme cases, the classical correlation parameters lose their real significance.

The scatter coefficient and the coefficients of linear importance always preserve a sense. These notions will be employed here for the discussion and classification of the various cases of linear dependency and the corresponding types of clustering. This analysis of types of clustering does not, of course, in itself exhaust the subject of correlation. What it does, is to furnish a logical and, in my mind, necessary supplement to the classical methods.

In the present section the rigorous algebraic point of view will be adopted. The statistical interpretation of the algebraic criteria is discussed in the following section. The methods to be used are rather simple. In fact, they are only based on the elementary properties of *Gram*-ian matrices. This is another reason why a rigorous analysis of linear dependencies and the corresponding types of clustering should not be neglected in the study of correlation between statistical variables. From the discussion in this section and the next it will immediately follow in which cases

the definition of the classical correlation parameters preserves a sense.

Supposing the variables to be measured from their means, we lay down the following definitions.

The n dimensional set of observational variables $x_1 \dots x_n$ is said to be linearly dependent, or to be a *collinear* set if there exists identically in t at least one linear relation of the form

$$a_1x_1(t)+\dots+a_nx_n(t)=0 \tag{6.1}$$

where the coefficients $a_1 \dots a_n$ are independent of t and not all equal to zero. If no such relation exists, the set is called linearly independent or *non collinear*.

If there exists identically in t exactly p $(0 \leqq p \leqq n)$ distinct linear relations of the form (6.1), the set is said to be *p-fold collinear* or *p-fold flattened*. By p distinct linear relations of the form (6.1) is meant p linear relations, such that not one of them can be derived as a linear combination of the others, with constant coefficients.

From the theory of linear equations it follows that this definition of a p-fold collinear set is equivalent whith the following: An n dimensional set is p-fold collinear when and only when there exists at least one $(n-p)$ dimensional subset $x_\mu \dots x_\vartheta$ which is non collinear and such that each of the remaining p variables, identically in t, can be expressed as a linear combination of $x_\mu \dots x_\vartheta$ with constant coefficients. Evidently, if the remaining p variables can be expressed as a linear combination of $x_\mu \dots x_\vartheta$ with constant coefficients, any of the n variables $x_1 \dots x_n$ can. A p-fold collinear n dimensional set is therefore a set, which by a non singular linear transformation can be transformed into a set where p of the variables are ineffective and $\varrho = n-p$ of the variables are effective.

Geometrically interpreted a p-fold collinear n dimensional set is a set for which the swarm of observation points in n dimensional space are (rigorously) crowded in a certain $\varrho = n-p$ dimensional plane through the origin (but not in a lower dimensional plane). This ϱ dimensional plane is called the *perfect regression plane* for the set, $\varrho = n-p$ is called the *rank* or the *unfolding capacity* of the set, p is called the *flattening* of the set.

If $p=0$, that is if the rank of the set is equal to its dimensionality, the observation points are scattered in n dimensional space. There is no flattening, and the set is non collinear.

If $p=1$, that is if there exists exactly one linear relation of the form (6.1), the set is called *simply collinear*. In this case the flattening is 1, and the rank is exactly one less than the dimensionality of the set. There exists a perfect $(n-1)$ dimensional regression plane.

If $p>1$, the set is called *multiply collinear* or *multiply flattened*. There now exists a perfect regression plane of lower dimensionality than $(n-1)$.

A simply collinear n dimensional set $x_1 \dots x_n$ is called a *closed* set if all the n coefficients $a_1 \dots a_n$ in the linear relation which holds good for the set, are different from zero. This definition of a closed set is equivalent with the following: An n dimensional set is closed if there exists at least one relation of the form (6.1) involving all the n variables (i. e. all the coefficients $a_1 \dots a_n \neq 0$) and no relation of the form (6.1) involving less than n variables. In fact, from this definition follows that the set must be simply collinear. For if there exist two (or more) distinct relations of the form (6.1), we can eliminate one variable and obtain a relation involving at most $(n-1)$ variables.

Geometrically interpreted a closed n dimensional set $x_1 \dots x_n$ is a simply collinear set, the perfect $(n-1)$ dimensional regression plane of which does not contain any of the axes $x_1 \dots x_n$. In this case each one of the variables in the set can be expressed in terms of the others.

I shall now consider necessary and sufficient criteria for the cases defined above. I consider the n dimensional set $\mathbf{x}=(x_1 \dots x_n)$, but the argument is identical for any q dimensional subset $x_\alpha \dots x_\gamma$, the q rowed matrices $\mathbf{M}_{q[\alpha \dots \gamma]}$ and $\mathbf{R}_{q[\alpha \dots \gamma]}$ then taking the place of the n rowed matrices $\mathbf{M}$ and $\mathbf{R}$.

I shall assume the set $\mathbf{x}$ to be an effective set, i. e. none of the variables is identically zero. In this case each minor contained in $\mathbf{M}$ is equal to or different from zero according as the corresponding minor of $\mathbf{R}$ is equal to or different from zero. In particular $\mathbf{M}$ and $\mathbf{R}$ have the same rank.

From the purely algebraic point of view one is only concerned with the cases where the set $\mathbf{x}$ has rigorously the property of being collinear, closed etc. Criteria for these cases may be derived either from the properties of $\mathbf{M}$ or from the properties of $\mathbf{R}$. From the statistical point of view (which will be discussed in

the next section) we have however to take account also of cases where the set $\mathbf{x}$ only "comes near" to having the properties in question, the sense of "coming near to" being defined by the values (close to zero or close to unity) of the determinant $R=|\mathbf{R}|$ and its minors. It is therefore preferable to state also the rigorous algebraic propositions in terms of the properties of $\mathbf{R}$. It should be noticed that in order to ascertain the rank of $\mathbf{R}$ it is sufficient to inspect its *principal* minors, for $\mathbf{R}$ is symmetric.

We have already seen that the set $\mathbf{x}$ is collinear when and only when $R=0$. A generalization of this is the following proposition: *The observational set* $\mathbf{x}$ *is of rank* ϱ *(i. e. its flattening is* $p=n-\varrho$*) when and only when the correlation matrix* $\mathbf{R}$ *is of rank* ϱ.

For if the set is of rank ϱ, there exist identically in t exactly $p=n-\varrho$ distinct linear relations between the variables $x_1 \dots x_n$. Therefore, given any $(\varrho+1)$ dimensional subset, we can always by elimination obtain a linear relation which involves at most these $(\varrho+1)$ variables, and which is satisfied identically in t. All the $(\varrho+1)$ rowed determinants of the form

$$\begin{vmatrix} x_\alpha(t_1) & \dots & x_\alpha(t_{\varrho+1}) \\ \dots & \dots & \dots \\ x_\gamma(t_1) & \dots & x_\gamma(t_{\varrho+1}) \end{vmatrix} \tag{6.2}$$

where $\alpha \dots \gamma$ are $(\varrho+1)$ arbitrary numbers from the sequence $1,2 \dots n$ must therefore vanish identically in $t_1 \dots t_{\varrho+1}$. And the same holds good a fortiori for the higher rowed determinants of this form. Hence by (1.2) all $(\varrho+1)$ and higher rowed principal minors of $\mathbf{R}$ must vanish. Furthermore, if the set $\mathbf{x}$ is of rank ϱ, there exists at least one ϱ dimensional subset $x_\mu \dots x_\vartheta$ for which no linear relation holds good identically in t. The determinant

$$\begin{vmatrix} x_\mu(t_1) & \dots & x_\mu(t_\varrho) \\ \dots & \dots & \dots \\ x_\vartheta(t_1) & \dots & x_\vartheta(t_\varrho) \end{vmatrix} \tag{6.3}$$

can therefore not vanish identically in $t_1 \dots t_\varrho$. There must be at least one set of values $t_1' \dots t_\varrho'$ for which (6.3) is different from zero. The sum square of (6.3) must therefore be positive, not zero, if the summation subscripts $t_1 \dots t_\varrho$ independently of each other run through all values of t. Hence the ϱ rowed principal minor $R_{\varrho[\mu \dots \vartheta]}$ in $\mathbf{R}$ must be different from zero. Therefore, $\mathbf{R}$ being symmetric, must be of the rank ϱ.

Inversely, if the symmetric matrix $\mathbf{R}$ is of rank ϱ, it must contain at least one ϱ rowed principal minor $R_{\varrho[\mu \dots \vartheta]}$ different from zero. Hence the set $x_\mu \dots x_\vartheta$ is linearly independent. There must consequently exist at least one set of values $t_1' \dots t_\varrho'$ for which (6.3) has a value a different from zero. But all $(\varrho+1)$ rowed determinants of the form (6.2) must vanish identically in $t_1 \dots t_{\varrho+1}$. In particular the $(\varrho+1)$rowed determinant

$$\begin{vmatrix} x_\mu(t_1') & \dots & x_\mu(t_\varrho) x_\mu(t) \\ \dots & \dots & \dots \\ x_\vartheta(t_1') & \dots & x_\vartheta(t_\varrho') x_\vartheta(t) \\ x_i(t_1') & \dots & x_i(t_\varrho) x_i(t) \end{vmatrix}$$

where i is any of the numbers $1,2 \dots n$, must vanish identically in t. Developping this determinant after the last column we see that there must exist identically in t a linear relation

$$a_{i\mu} x_\mu(t) + \dots + a_{i\vartheta} x_\vartheta(t) + a x_i(t) = 0$$

where the coefficients are independent of t, and at least one of them, namely a, is different from zero. Incidentally, since all the variables are assumed effective, there must even be at least two coefficients in the last equation, which are different from zero. If $\mathbf{R}$ is of rank ϱ, each of the variables in the set $\mathbf{x}$ can therefore, identically in t, be expressed linearly in terms of a certain linearly independent ϱ dimensional subset. Hence the set $\mathbf{x}$ is of rank ϱ.

In order to find a linearly independent ϱ dimensional subset contained in the set $\mathbf{x}$, we have to inspect the ϱ rowed principal minors of $\mathbf{R}$. Any subset $x_\mu \dots x_\vartheta$ such that $R_{\varrho[\mu \dots \vartheta]}$ is different from zero, is a linearly independent set and can be used for expressing all the variables in the set $\mathbf{x}$.

If the number n of variables is equal to number ω of observations, the set is always collinear. More generally: If $\omega \leqq n$, the rank of the set is at most equal to $\omega-1$, i. e. its flattening is at least equal to $n-\omega+1$. This simply follows from the fact that ω points $\mathbf{x}(1) \dots \mathbf{x}(\omega)$ between which there exists the linear relation $\Sigma_t \mathbf{x}(t)=0$ (or any homegeneous linear relation), must necessarily lie in an $\omega-1$ dimensional plane through the origin (and they might even lie in a lower dimensional plane).

This fact is also revealed by the rank of $\mathbf{R}$. The general term

$$(6.4) \qquad \begin{vmatrix} x_\alpha(t_1) \dots x_\alpha(t_\omega) \\ \dots\dots\dots\dots \\ x_\gamma(t_1) \dots x_\gamma(t_\omega) \end{vmatrix}^2$$

in the formula for the ω rowed principal minor $R_{\omega[\alpha \dots \gamma]}$ in $\mathbf{R}$ is namely zero if two or more of the t_i are equal. And if none of t_i are equal, the set of values $t_1 \dots t_\omega$ must be exactly the set of ω different values which represent the ω observations. Adding to any column in (6.4) all the other columns we therefore get a determinant where one column consists exclusively of zeros (since the variables are measured from their means). Hence all the ω rowed principal minors of $\mathbf{R}$ are zero. That all the higher rowed principal minors of $\mathbf{R}$ must be zero, simply follows from the fact that in the expression for any of these principal minors each term is a determinant with at least two equal columns.

From the preceding proposition we immediately infer the following: *A necessary and sufficient condition for a set* $\mathbf{x}$ *to be simply collinear is that* $R=0$ *and that at least one of the diagonal elements* $\hat{r}_{ii}$ *in the adjoint correlation matrix* $\hat{\mathbf{R}}$ *is different from zero.*

If an n dimensional set is simply collinear, it is always possible, at least in one way, to pick out a subset of $(n-1)$ variables in terms of which the one remaining variable can be expressed linearly. But it is by no means certain that *any* of the variables can thus be expressed in terms of the others. *The variable* x_i *in a simply collinear set can be expressed linearly in terms of the other variables when and only when* $\hat{r}_{ii} \neq 0$. For x_i can evidently be expressed in terms of the other variables when and only when x_i occurs in the one existing relation which holds good for the simply collinear set, that is, when and only when the remaining $(n-1)$ variables do not by themselves from a collinear set, i. e. when and only when $\hat{r}_{ii} \neq 0$. This fact can also be inferred from (4. 23).

From this we further conclude: *A necessary and sufficient condition for a set* $\mathbf{x}$ *to be a closed set is that* $R=0$ *and all the diagonal elements* $\hat{r}_{ii}$ *of the adjoint correlation matrix are different from zero.*

In a simply collinear set, a given row (column) of the adjoint moment matrix $\hat{\mathbf{M}}$ will consist exclusively of zeros when and only when the diagonal element in this row (column) is equal to zero. For if $\hat{\mathbf{M}}$ is of rank 1, we have $\hat{m}_{ij}^2 = \hat{m}_{ii}\hat{m}_{jj}$ $(i.j=1,2 \dots n)$, where at least one of the $\hat{m}_{ii}$ is $\neq 0$. The same rule holds good for the adjoint correlation matrix.

The least relation also shows that in a simply collinear set, a given element $\hat{r}_{ij}$ of $\hat{\mathbf{R}}$ will vanish when and only when at least one of the two quantities $\hat{r}_{ii}$ and $\hat{r}_{jj}$ vanishes. If either $\hat{r}_{ii}$ or $\hat{r}_{jj}$ (or both) vanishes, the partial correlation coefficient $\overline{r}_{ij}$ will consequently be of the undefined 0/0 form. If $\hat{r}_{ii}$ vanishes, the multiple correlation coefficient r_i will be of the form 0/0. Hence: *If all the partial (or all the multiple) correlation coefficients in a simply collinear set shall have a meaning, the set must be a closed set.* From (4.25) and (4.27) is seen that in a closed set all the $\overline{r}_{ij}$ and all the r_i have an absolute value equal to unity. The case of a closed set might therefore also be designated as the case of perfect correlation.

If the simply collinear set is not closed, all those partial correlation coefficients $\overline{r}_{ij}$ will be undefined which are situated in $\overline{\mathbf{R}}$ in a row (or a column) whose corresponding diagonal element in $\hat{\mathbf{R}}$ is zero.

If the set is multiply collinear, the indeterminateness is even greater. In this case $\hat{\mathbf{R}}$ is of rank 0, and therefore each and all of the partial and multiple correlation coefficients of highest order undefined.

7. TYPES OF CLUSTERING AND NON SIGNIFICANT CORRELATIONS.

The discussion of the present section will concern the statistical interpretation of the notion of types of clustering, and particulary the significance of this notion in the procedure of fitting lines and planes to a given scatter diagram. The point of view which will be adopted is thus the regression point of view as distinguished from the frequency distribution point of view, although one or two occasional references will be made to the frequency aspect of the problem.

The procedure of determining an analytic relation between a certain number of statistical variables, has a conceptual background somewhat like this. We have in our mind a certain pattern, according to which variations in statistical variables can be of three kinds. *Accidental* variations are variations due to the fact that a great number of variables has been overlooked, consciously or unconsciously, each of the variables being however of minor importance. *Disturbances* are variations due to the fact that one single, or a certain limited number of highly significant variables have been overlooked. *Systematic variations* are variations which show a certain regularity (in time or space), this regularity being

taken as a criterion that no really relevant variable has been overlooked.

This classification does not pretend to be anything like satisfactory from a philosophical point of view. A critical reader will for instance find the troublesome problem of causality hidden in practically every line of the definitions. Furthermore the distinction between what shall be considered as accidental variations and disturbances, is not sharp. To some extent the distinction depends on the complexity of the problem and on the relative perfection (or deficiency) of the empirical and rational tools of investigation which are at present at our disposal. Thus, in economics we are actually often forced to throw so much into the bag of accidental variations that this kind of variations comes very near to take on the character of disturbances. In such cases it would perhaps be more rational to introduce an hierarchic order of types of variations, each type corresponding to the overlooking of variables of a certain order of importance.

Nevertheless, I think the rough three-fold distinction which has been made above, will be sufficient for our purpose, so that it shall not be necessary to enter upon a lengthy philosophical discussion.

When we proceed to the determination of an analytic relation between a certain number of statistical variables, we assume, implicitly or explicitly, that if the considered set of variables really forms a complete set of relevant variables, and if the accidental variations could be eliminated, so that only the systematic variations were left, then the variables would satisfy a certain functional relation (anyhow for some limited time or space considered). In fact, this assumption is really involved in the definition of systematic variations as distinguished from accidental variations and disturbances. The character of this functional relation is an important feature in any statistical problem, and in many problems it is the one important thing in which we are interested. A statistical fitting procedure, performed with a tentative analytic formula, is an attempt to get rid of the non-significant accidental variations and thus obtain some idea of the character of the functional relation which exists between the systematic variations.

As I see it, any statistical fitting procedure can be considered from this point of view. This interpretation seems rather natural. It is not, however, the only one which has been advanced. The

set of n elementary regressions (4.7) is sometimes considered as a unity, and as such contrasted with the unique regression obtained by some kind of mean regression method. It is contended that in principle only a regression of the latter type represents an approximation to the functional relation between the systematic variations, and that the set of n elementary regressions represents an entirely different notion, namely the notion of stochastic relation, the stochastic relation being not a unique relation such as the functional relation, but a plurality of relations, namely as many relations as there are variables. It is claimed that this distinction is fundamental and characteristic for the distinction between the mechanical and the statistical conception of "law".

In my mind this interpretation is fallacious. As pointed out in Section 4, the difference between the various conceivable regressions is a difference in assumption as to *how* the accidental element has actually manifested its presence in the material at hand, this difference in assumption entailing a difference in the technique by which the regression coefficients are determined. In the case of a mean regression the assumption and the technique is more symmetric in the variables than in the case of the elementary regressions. In a certain type of problems the technically best approximation to the functional relation between the variables will therefore be furnished by a mean regression, in another type of problems it will be furnished by one particular of the elementary regressions. It is only in this technical sense that a mean regression is "mean". Otherwise there is no difference between a mean regression and one particular of the elementary regressions. And it does not seen plausible to pick out in the infinity of possible techniques, that particular kind of technique which leads to the system of elementary regressions, and erect it into a principle, the conceptual importance of which should be comparable with the basic importance of the idea of functional relationship. There certainly does exist a difference between the conceptual schemes of a mechanical and a statistical law, but not in the sense that the first is something unique, the second something pluralistic. The difference, as I see it, is that the first is a law conceived so as to admit of no exceptions, the second a law which really admits of exceptions, just because the accidental variations are thought of as being superimposed on the systematic variations.

Another point which should be noticed in connection with the conception of regression and functional relation, is that the partial and multiple correlation coefficients (and the generalizations of these parameters to curvilinear regression) are not primarily descriptive of the character of the systematic variations, but are essentially indicators of the presence of accidental variations and disturbances. In the exaggerated importance which in recent years has been attributed to the computation of partial and multiple correlation coefficients, one has lost of sight to some extent, it seems to me, the fact that one of the essential things we are after, is the character of the regressional relation itself.

After these preliminary remarks we may turn to the notion of types of clustering. Let a statistical material covering ω observations on n variables be given. And suppose we want to investigate the character of the systematic variations in the variables, by assuming as a tentative analytic formula the linear function.

In all practical cases the moment matrix, and hence the correlation matrix will be non singular. This in itself does not, however, tell us very much from the statistical point of view. The essential question from the statistical point of view is if the deviation of R from zero is significant or not, that is, if the deviation of R from zero is really descriptive of the systematic variations in the variables. Even if the systematic variations of the variables are such that they would give rigorously $R=0$, *the slightest amount of accidental variation introduced would at once make* R *positive, not zero.* If R is considered from the sampling point of view, assuming the ideal universe in question to have R rigorously equal to zero, the mathematical expectation of an actually observed R would not be zero but some positive quantity. And the probability of an observed $R=0$ would be virtually zero. An actually observed value of R has therefore to be viewed in the light of some criterion of the significance of its deviation from zero. And similarly for the other scatter parameters, that is, for the principal minors of R, respectively the square roots of R and its principal minors.

For a rigorous analysis it would be highly desirable to have an exact criterion for the significance of the observed magnitude of the scatter parameters, in the form of formulae for the mathematical expection and standard deviations on these quantities, or better still: in the form of complete theoretical distributions. At present I have no such formulae to offer. But nevertheless we

have a rough criterion by which to judge the scatter parameters closeness to zero, namely the fact that all these quantities are lying between 0 and 1 and are limited by the formulae and propositions of Section 2. We are thus virtually in the same position with regard to judging the magnitude of the scatter parameters as we are with regard to judging the magnitude of those classical correlation parameters for which mathematical expectations, standard deviations or complete theoretical distributions are not available at present. This sort of criterion is certainly not ideal. But if it is used with care, I believe it is far better than no criterion at all. And it is perhaps not vain to hope that it shall ultimately be possible to derive the necessary formulae for giving a more definite meaning to the scatter parameters closeness to zero.

In view of the algebraic facts indicated in Section 6 it is clear that the notions of simply collinear sets and closed sets must be important from the statistical point of view.

Suppose for instance that the scatter coefficient $s=\sqrt{R}$ for an n dimensional set is found to be not significantly different from zero, indicating that the observation points in the n dimensional scatter diagram come close to lying in a plane. This in itself is not sufficient to make it a plausible procedure to pick out one of the variables, say x_i, and compute the regression of x_i on the remaining $(n-1)$ variables. This procedure would have no meaning if $\sqrt{\hat{r}_{ii}}$ is not significantly different from zero, while at least one other diagonal element in $\hat{\mathbf{R}}$ is significantly different from zero. For in this case the $(n-1)$ dimensional plane around which the observation points are clustering is *not significantly different from a plane which contains the* x_i *axis*, the points being highly scattered within this plane. The regression which would have a meaning would therefore be a regression between (all or some of) the remaining $(n-1)$ variables, *not a regression involving* x_i. The variable x_i has nothing to do in the linear regression system. From the point of view of linear regression x_i is a superfluous variable drawn into observation. The n dimensional set is not a closed set. If we would compute the regression of x_i on the other variables in this case, the whole system of regression coefficients b_{ij} would be artificial. Computing the regression of x_i on the other variables would namely now mean forcing the quantity $\sqrt{\hat{r}_{ii}}$ (whose deviation from zero is non-significant) into the coefficients b_{ij} as a denominator.

The question may arise: Is not this exactly the kind of thing which would show up in the multiple and partial correlation coefficients? If the actual distribution of the observations is such that the variable x_i is a superfluous variable in the above sense, would not the partial correlations $\overline{r}_{ij}$ and the multiple correlation r_i necessarily have small numerical values? The answer is: They would not. On the contrary, the definition of the partial and multiple correlation coefficients is such that when the remaining $(n-1)$ variables taken by themselves come close to forming a collinear set, the partial correlations $\overline{r}_{ij}$ between the superfluous variable x_i and any of the remaining variables, respectively the multiple correlation r_i between x_i and the set of the other $(n-1)$ variables, may assume any value, in particular these parameters might come close to unity, *thus making it appear perfectly legitimate to compute the regression of x_i on the other variables.*

I shall take the case $n=3$ as an illustration. Let r_{12}, r_{13} and r_{23} be the simple correlation coefficients in the set $(x_1x_2x_3)$. We put $r_{12}=ph$, $r_{13}=qh$ and $r_{23}=\varepsilon\sqrt{1-h^2}$ $(h^2 \gtreqless 1)$, where ε is the sign of r_{23} and the square root is taken positive. Now consider p, q and h as arbitrary quantities. The consistency condition which characterizes the case where r_{12}, r_{13} and r_{23} are correlation coefficients for a set of real variables is

$$0 \gtreqless R = h^2(1-[p^2-\varepsilon\cdot 2pq\sqrt{1-h^2}+q^2]).$$

By studying the two conics in (p, q)-coordinates

$$p^2-\varepsilon\cdot 2pq\sqrt{1-h^2}+q^2=1$$
$$\varepsilon=\pm 1$$

which have the shape indicated in Fig. 2, we see that the consistency condition is certainly satisfied for all values of $h^2 \gtreqless 1$ if the point (p,q) is situated in the inner square of the figure, that is if

$$|p|+|q| \gtreqless 1. \tag{7.1}$$

In terms of p,q and h we have (using Professor *Yule*'s notation for the partial and multiple correlation coefficients)

$$r_{12\cdot 3}=(p-\varepsilon q\sqrt{1-h^2})/\sqrt{1-q^2h^2}$$
$$r_{13\cdot 2}=(q-\varepsilon p\sqrt{1-h^2})/\sqrt{1-p^2h^2}$$
$$R_{1(23)}=\sqrt{p^2-\varepsilon\cdot 2pq\sqrt{1-h^2}+q^2}$$

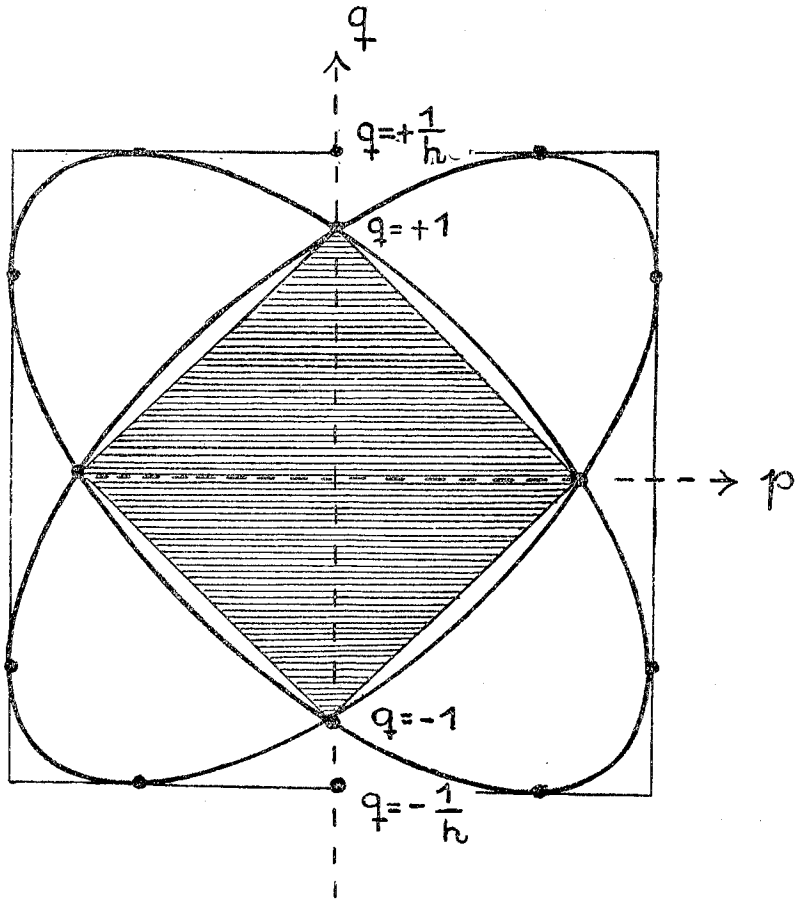

Fig. 2.

where all the square roots are taken positive. Further the standard error of estimate of x_1 as a linear function of x_2 and x_3, is

$$\sigma_{1\cdot 23}=\sigma_1\sqrt{1-R^2_{1(23)}}$$

where σ_1 is the simple standard deviation of x_1.

Therefore, if p and q are chosen as two arbitrary numbers independent of h and satisfying (7.1), we have at the limit for $h \to 0$.

$$r_{12\cdot 3}=p-\varepsilon q$$
$$r_{13\cdot 2}=q-\varepsilon p$$
$$R_{1(23)}=|p-\varepsilon q|.$$

Now, the limiting process $h \to 0$ means that we construct a case where the observation points in the three dimensional scatter diagram $(x_1x_2x_3)$ come close to lying in a plane containing the x_1-axis (i. e. a plane perpendicular to the x_2x_3 plane). But in this plane the observation points are highly scattered. The observation points are far from clustering around a straight line in this plane. In other words, we construct a case where the set $(x_1x_2x_3)$ comes as near as we want to being a simply collinear but not a closed set, namely a set where there exists a linear relation between x_2 and x_3 taken by themselves, but where there is no approach whatsoever to a linear relation between x_1 and the other variables. The case is one where x_1 is a superfluous variable drawn into observation.

Nevertheless we can specify the case by disposing of p and q in such a way that for a sufficientlv small h any of the partial correlation coefficients between the superfluous variable x_1 and one of the other variables, for instance the partial correlation coefficient $r_{12\cdot3}$ comes as near as we please to any number between -1 and $+1$. Or the multiple correlation coefficient $R_{1(23)}$ can be brought as near as we please to any number between 0 and 1. For r_{23} positive we may for instance choose $p = -q = \frac{1}{2}$. At the limit for $h \to 0$ we get $r_{12\cdot3} = +1$, $r_{13\cdot2} = -1$, $R_{1(23)} = 1$ and $\sigma_{1\cdot23} = 0$. Choosing $p = q = \frac{1}{2}$ we get $r_{12\cdot3} = r_{13\cdot2} = 0$ etc.

The case $p = -q = \frac{1}{2}$ is particularly illustrative. All criteria seem to indicate that it would be perfectly legitimate to compute the regression of x_1 on the two other variables. We have maximum partial and multiple correlation and minimum standard error of estimate. And still such a regression would have no sense.

If we had computed the scatter coefficient s and the coefficients of linear importance

$$s = +\sqrt{R} = h\sqrt{1-(p^2-\varepsilon\cdot 2pq\sqrt{1-h^2}+q^2)}$$

$$\sqrt{\hat{r}_{11}} = h$$

$$\sqrt{\hat{r}_{22}} = \sqrt{1-q^2h^2}$$

$$\sqrt{\hat{r}_{33}} = \sqrt{1-p^2h^2}$$

where all the square roots are taken positive, the real character of the relation between the variables would at once have been revealed, for we have at the limit for $h \to 0$

$$s = \sqrt{\hat{r}_{11}} = 0 \text{ and } \sqrt{\hat{r}_{22}} = \sqrt{\hat{r}_{33}} = 1$$

$s = 0$ means that there exists at least one linear relation between the three variables, $\sqrt{\hat{r}_{22}} = 1$ (or $\sqrt{\hat{r}_{33}} = 1$) means that there does not exist more than one such relation, and $\sqrt{\hat{r}_{11}} = 0$ means that the one existing relation is a relation between x_2 and x_3.

It is easy to construct numerical examples which illustrate this type of clustering. I constructed for instance three series $x_1x_2x_3$ where x_1 was a combination of words in english, each letter being attributed a value according to its place in the alphabet, x_2 consisted of digits picked at random in a logarithmic table, and x_3 was nearly a linear function of x_2. This case happened to give a high negative value for $r_{12\cdot3}$ and high positive values for $r_{13\cdot2}$ and $R_{1(23)}$. But the value of $\sqrt{\hat{r}_{11}}$ (and therefore necessarily the value of s) turned out to be comparatively small, while $\sqrt{\hat{r}_{22}}$ and $\sqrt{\hat{r}_{33}}$ were significantly different from zero, indicating the presence of a linear relation between x_2 and x_3 and the lack of a linear relation involving x_1.

The conclusion is that before proceeding to the computation of the classical correlation parameters, it will be advisable to take a general survey of the type of clustering, using the scatter coefficient and the coefficients of linear importance, and if necessary the scatter coefficients for the lower dimensional subsets.

In particular it is essential to determine if the set (assumed approximately collinear) comes near to being a closed set, and if not to pick out those subsets which come near to being closed. It is only for the approximately closed subsets thus determined that the classical correlation parameters have a real significance. In particular it is only for a rigorously closed set the term perfect correlation has a meaninng.

The following might serve as a general scheme for the analysis. First compute the simple correlation coefficients, i. e. the elements of the correlation matrix $\mathbf{R}$. If the scatter coefficient $s = +\sqrt{R}$, where $R = |\mathbf{R}|$, is close to unity, there is no use trying to express any of the variables linearly in terms of the others.

If s is reasonably close to zero, the diagonal elements $\hat{r}_{ii}$ in the adjoint correlation matrix should be computed. If none of the quantities $\sqrt{\hat{r}_{ii}}$ are very small, the set may be considered as a closed set and if desired the regression coefficients and other

classical correlation parameters computed in the usual way. If a mean regression is wanted, one of the forms (4.16), (4.22) or (5.22) may be tried.

If some of (but not all) the quantities $\sqrt{\hat{r}_{ii}}$ are very small, the set may still be considered as a simply collinear but no longer as a closed set. Those variables x_α for which $\sqrt{\hat{r}_{\alpha\alpha}}$ is very small, might be left out and the rest of the variables treated as a closed set. If it is desired not to leave the variables x_α out completely, one might compute the regressions in the usual way, however not compute the regression of any of the x_α on the other variables.

If all the quantities $\sqrt{\hat{r}_{ii}}$ are very small, the set should be considered as multiply collinear. In this case there will exist at least two subsets which may be considered as closed sets and treated separately. These closed subsets are determined by an inspection of the $(n-2)$ rowed and eventually the lower rowed principal minors of the correlation matrix **R**. If there exists at least one ϱ rowed principal minor of **R**, the positive square root of which is not a very small quantity, while the positive square roots of all the higher rowed principal minors of **R** are very small, then the given n dimensional set should be considered as $p=n-\varrho$ fold collinear (p-fold flattened). There now exist exactly p closed subsets which may be treated separately. These p subsets are determined in the following way. Pick out the ϱ dimensional subset $x_\mu \dots x_\delta$ which is such that $R_{\varrho[\mu \dots \delta]}$ is the greatest of all ϱ rowed principal minors in **R**. This ϱ dimensional subset might be called the *basis set*. The basis set is the ϱ dimensional subset which comes nearest to being an uncorrelated set. Now consider in turn the p $(\varrho+1)$ dimensional subsets obtained by adding to the basis set one of the variables which are not in the basis set. Each of these p subsets comes close to being a simply collinear set and might be analysed as such. In particular the set might be reduced to a closed set by omitting all those variables x_α which are such that the square root of the corresponding diagonal element in the $(\varrho+1)$ rowed adjoint correlation matrix for the subset, is a very small quantity. The variables which will eventually be omitted by this rule, are necessarily variables occuring in the basis set. The one variable which is added to the basis set, will never be omitted.

RAGNAR FRISCH

pro tem New York City, February 1928

INTERPOLATION I STATISTIKEN.

Udgiverens Ønske om et Bidrag vedrørende Interpolation i Statistiken har jeg opfattet som et Ønske om saadanne Bemærkninger, der vil ligge den praktisk udøvende Statistiker nærmest; ogsaa paa Grund af den forholdsvis begrænsede Tid, der har været stillet til min Raadighed, har jeg mere maattet lægge Vægten paa den praktiske end paa den teoretiske Side af Sagen og derfor mere maattet søge at give Eksempler paa, hvorledes man kan bære sig ad overfor de Vanskeligheder, i hvilke Statistiken paakalder interpolatorisk Hjælp, end paa at undersøge, i hvilken Udstrækning Metoderne kan tænkes at have Gyldighed eller Værdi.

Det er nu ganske vist mere en Talemaade end en Realitet, naar man i denne Forbindelse fremhæver den praktiske Side paa Teoriens Bekostning eller omvendt. I Virkeligheden lader disse to Sider sig ikke adskille; de følgende Bemærkninger vil formentlig give et Indtryk af Hensigten med denne Adskillelse.

1.

Indskrænker vi os foreløbigt til den i Formen simpleste Interpolationsopgave, nemlig ud fra Kendskabet til visse (i Reglen forholdsvis faa) Funktionsværdier (Tabelværdierne) at beregne andre (muligvis vilkaarlig mange), maa det som karakteristisk for denne Opgave fremhæves, at det i Hovedsagen skal være *ud fra de givne Tabelværdier* der drages Slutning om Funktionens Værdi for andre Argumenter, hvorimod en Beregning, der direkte følger den Vej, Funktionens Definition anviser, ikke kommer i Betragtning som "en Interpolation». Ved tilstrækkelig Ihærdighed kan man eksempelvis beregne Værdien af $\sqrt[3]{5\frac{1}{2}}$ med forud forlangt Nøjagtighed alene ved den Regningsart (Potensopløftning), hvorved Roduddragning defineres, og det er derved ganske upaakrævet at kende Værdierne f. Eks. af $\sqrt[3]{4}$, $\sqrt[3]{5}$, $\sqrt[3]{6}$ og $\sqrt[3]{7}$. Kender man imidlertid Værdien af disse fire Størrelser, er det, at man uden for-